Building Reliable Credit Rating System for Firms in China

By
BAI Ling

A Thesis Submitted in Partial Fulfillment
of the Requirements for the Degree of
Master of Philosophy
in
Systems Engineering and Engineering Management

ABSTRACT

After China joined the WTO and the New Basel Capital Accord - Basel II introduced to China, better credit risk management has become extremely important to domestic banks. By 2007, China's financial market will be opened to foreign competitors and domestic banks will face more fierce competitions. One of the important issues faced by domestic banks is how to design and develop a sound credit rating system that addressed a particular issue in China – inadequate efforts to guarantee disclosed financial data truly reflect companies' financial and operating performance. Although credit rating is well developed in the Western countries, the methodologies and applications are insufficient to be directly applied to measure credit worthiness of firms in China because the due-diligence problem cannot be solved. It is believed that unless such "falsified" or "untrue" components were removed, the results produced would be meaningless or misleading. With above considerations, we proposed a study framework named COMPLETE as a guideline to evaluate credit risks for firms in China. The COMPLETE framework includes 8 aspects, namely **Capital, Operation, Management, Political, Leverage, Earning, Trustworthiness** and **Environment**, where **Trustworthiness** is dedicated to tackle the due-diligence problem. We also proposed three prototypes of rating process. The first proposal included a subsystem called **continuous monitoring,** and it is designed to avoid financial statement frauds by strengthening regulatory enforcements. The latter two proposals were aimed at developing quantitative models to evaluate financial data reliability and integrate the result to credit risk assessments.

Multivariate statistical techniques were employed to find key bankruptcy indicators and develop the credit scoring model. The analysis over 72 A-share companies listed in Shanghai and Shenzhen exchange indicated that variables related to leverage, earnings, short-term liquidity, sustainability of development, asset utilization and ability of capital accumulation could provide helpful suggestions for credit analysts. A credit scoring model was developed with Logit regression, and it achieved an overall prediction accuracy of 90.28%. The impact of false financial statements (FFS) was investigated through a case study approach and a quantitative approach. An in depth analysis of Hubei Lantian Co. Ltd.'s financial scandal case was conducted first, and we demonstrated how to apply the COMPLETE framework to evaluate the trustworthiness aspect of a firm. Then, a formal FFS analysis was conducted using Classification and Regression Tree (CART) as the methodology. It is found that CART is effective in discriminating FFS from non-FFS. Other interesting findings include a comparison of seriousness of FFS between China and Western firms, and manipulations in Chinese firms were usually made on a consecutive basis. These findings provided valuable suggestions for building up a reliable credit rating system.

论文摘要

中国加入世贸，新巴塞尔协议引入中国，这一切都促使中国银行业改善其风险管理机制。到 2007 年，中国的金融市场将全面开放，那时国内的银行将面临更严峻的竞争压力。有效的风险控管能帮助银行减少损失，提高投资者对其的信心，而其中的一项重要工作就是如何设计一套合理的信贷评级系统，来帮助银行评核其贷款者的信贷风险。在设计这样一个系统的时候，一个突出的问题就是如何判断贷款方的财务数据是否作假。虽说信贷风险的研究在西方发达国家已经趋于完善，但它们采用的方法几乎没有考虑到数据真实性的问题，因此将这些方法照搬到中国是不合理的。我们相信只有将这些数据去伪存真，才会得到准确公正的结果。

我们的信贷评级框架 COMPLETE 包括八个方面：资本质量，营运能力，管理机制，政策因素，杠杆效应，盈利能力，可信程度和环境因素。其中，可信程度是衡量财务报表水分的一个指标。我们设计了三种评级系统的模板，第一种包括了一个叫做“连续监控”的子系统，意在通过加强外部监管来阻止虚假财务数据的生成。另外两个模板基于对财务报表数据的量化分析。我们首先研究了破产风险的影响因素，通过对 72 家上市公司的单变量分析，我们发现杠杆效应，盈利能力，资产流动性，可持续发展能力，资产利用率和资产累积能力是几个重要的因素。我们用逻辑回归法（Logit Regression）建立了一个破产预测模型，用代入法检验的预测准确度达到了 90.28%。我们同时研究了假账对破产风险评估的影响，在此我们运用了案例分析和量化分析两种方法。我们首先运用信贷评级框架 COMPLETE 对蓝田股份管理舞弊案作了深入分析，我们相信整个分析过程能帮助我们了解中国企业的问题所在。我们进一步运用分类回归树（CART）的方法进行分析，我们发现回归树能有效的将虚假报表和真实报表区分开来。我们将分析结果同国外同类研究的结果进行了比较，我们也发现这些公司往往在上市之初就作假，而且在上一份假账的基础上继续制造新的假账。这些发现都有助于我们有效的鉴别财务数据的真伪，建立一个可靠的评级系统。

CONTENTS

INTRODUCTION 1
1.1 Background 1
1.2 Research Approach and Design 4
1.3 Organization of the Thesis 5
LITERATURE REVIEW ON CREDIT RISK MODELING 6
2.1 Overview 6
2.2 Discriminant Analysis 8
2.3 Logit Regression 10
2.4 Regression and Classification Tree (CART) 10
2.3 Chapter Summary 12
FALSE FINANCIAL STATEMENTS DETECTION 13
3.1 Overview 13
3.2 Empirical Studies on Financial Scandal and False Financial Statements (FFS) 14
3.3 False Financial Statements (FFS) Detection 17
3.4 Chapter Summary 23
RESEARCH METHODOLOGIES IN CREDIT SCORING & FALSE FINANCIAL STATEMENTS DETECTION 25
4.1 Overview 25
4.2 Logit Regression 26
4.3 Classification and Regression Tree (CART) 31
4.4 Chapter Summary 34
PROPOSED STUDY FRAMEWORK 35
5.1 The COMPLETE Framework 35
5.2 Rating Process 43
5.3 Chapter Summary 45
DEVELOPING THE CREDIT SCORING MODEL 46
6.1 Overview 46
6.2 Sample 46
6.3 Variables 47
6.4 Result of the Univariate Analysis 49
6.5 Develop the Bankruptcy Risk Model with Logit Regression 50
6.6 Chapter Summary 54
INVESTIGATING FALSE FINANCIAL STATEMENTS 55
7.1 Overview 55
7.2 Impact of False Financial Statements (FFS) on Credit Risk Assessments – Evidence from Lantian's Case 55
7.3 Evaluating the Trustworthiness Aspect for Lantian 56
7.4 Analyze FFS with Statistical Tools 59
7.5 Chapter Summary 73
SUMMARY 75
REFERENCES 77

CHAPTER 1
INTRODUCTION

1.1 Background

1.1.1 Research Significance and Objectives

Since China entered the WTO on 11-December 2001, a strong momentum has been witnessed in the growth and opening up of its banking industry. Huge impacts will be expected on the management of the risks that China's banks exposed. The Chinese government and the China Banking Regulatory Commission (CBRC) have been engaged in further opening up China's banking market to comply with the WTO principles and China's WTO commitments. For example, former restrictions on foreign banks in China are relaxed, and domestic banks will face more fierce competitions in the future. Even though the opening up process has been slow, at least we know that the banking industry will be fully opened by 2007 (Liu Mingkang 2003).

However, China's banking system still has a long way to improve. Under China's planned economy, a mono-banking system existed where each commercial bank also operated like a central bank: the government controlled interest rates and imposed credit plans or advised whom to lend the money to. Consequently, banks extended a large proportion of *policy loans* to State-owned Enterprises (SOEs) and constantly underwrite to loss-making enterprises. This situation did not improve fundamentally until the "big ban" enterprise reform package in 1997, when banking system was forced to undergo a major restructuring. The restructuring is focused on reducing *non-performing loans* (NPLs), preventing financial crises and limiting financial risks (Yuwa Wei 2003). However, Chinese banking sector still lags far behind its international counterparties due to heavy historical burdens, imperfect corporate governance, inefficient management, and weak innovation capacity. One of the evidence is that the average NPL rate of the four major State-owned banks was 23.5%, while the average number in foreign banks was maintained below 8% (CBRC 2004).

As the New Basel Capital Accord – Basel II has been introduced to China, it is expected to change domestic banks' way to identify, measure, and control risks. Although CBRC indicated that only Pillar 2 (Supervisory Review) and Pillar 3 (Market Discipline) would be adopted in the near future (Liu Mingkang 2003), they did emphasize that banks should improve their risk management practice. "All banks should start collecting the necessary data of their borrowers and facilities, which are the fundamental for a more quantitative approach to measuring and managing credit risk. We will consider the use of an *internal ratings-based* (IRB) approach, either foundation or advanced, for *minimum capital requirement* (CAR) calculation and regulation when banks are ready and also we should provide incentives for banks to improve their sophistication in risk management accordingly" (CBRC 2004). As a result, it is believed that developing a sound public rating system would help domestic banks in two folds: firstly, such information is one of the most important components in measuring credit risks in Basel II; besides, the experiences in developing such a system could provide valuable suggestions to domestic banks for building up more sophisticated or internal based risk management systems.

Huge number of researches had been conducted in credit rating and credit risk assessments for US, European, East Asian and South-east Asian firms (Beaver 1966, Altman 1968, Edmister 1972, Altman et al. 1977, Shirata 1998, Persons 1999, Crouhy et al. 2000, Jae H. Min & Youngchan Lee 2004), but few studies had been conducted on Chinese firms. Therefore, it is necessary to develop a credit risk model for firms in China. Two questions are concerned in this study: firstly, what are the indicators to reflect credit risk; secondly and more critically, how to make the credit rating robust to *highly unreliable financial data* disclosed by Chinese firms. Previous credit risk studies seldom touch this issue, and our research is aimed at combining due diligence treatment into credit risk assessment, and developing a reliable credit rating system.

1.1.2 Introduction to Credit Rating

Credit rating is emerged in the US in early 20^{th} century together with the issuance of corporate bonds. Today, the US remains its leading position in research and practice of credit rating. Two of the world's largest credit rating companies are S&P and Moody's, almost all public issued debt instruments in the US and Canada are rated by these two agencies; their ratings are also widely accepted by market participants and regulatory agencies. In general, a credit rating is not an investment recommendation concerning a given security. In the words of S&P, "a credit rating is S&P's opinion of the general credit worthiness of an obligor, or the credit worthiness of an obligor with respect to a particular debt security or other financial obligation, based on relevant risk factors." (Crouhy et al. 2000).

Credit rating systems are usually applied to non-financial corporations, as special approaches need to be employed for banks and financial institutions. There are different classifications for credit rating systems. Regarding the objective, there are *issuer credit ratings* and *issue-specific credit ratings*: issuer credit ratings is an opinion on the obligor's overall capacity to meet its financial obligations; issue-specific credit ratings are subject to particular debt instruments the company issued. Concerning the term to maturity, there are *long-term ratings* and *short-term ratings*: short-term ratings are applicable to commercial paper (CP), certificate of deposits (CD), and other short-term securities; long-term ratings usually apply to corporate bonds. The credit rating process includes both quantitative and qualitative analyses. The quantitative analysis is mainly financial analysis based on the firm's financial reports; the qualitative analysis also concerns the quality of management, the firm's competitiveness within its industry, the expected industry growth, its vulnerability to technological and regulatory changes (Crouhy et al. 2000). The Moody's rating analysis is illustrated in Figure 1.

Figure 1: Moody's rating analysis (Moody's, 2002)

The credit rating process typically starts from the issuer's request. Then the rating agency sends an analytical team to conduct basic researches. After that, a meeting with the management of the issuer is arranged and the credit agency reviews the issuer's operating and financial plans, policies and strategies. Once all the materials are collected, a rating committee with appropriate expertise in the relevant industry is formed; the materials are reviewed again by committee members and they vote for the recommendation. In the mean time, the issuer could appeal against the rating before it is announced. The final decision is usually issued 4 to 6 weeks after the request. The ratings are typically reviewed once a year based on new financial reports and business information. A "credit watch" or "rating review" notice is issued if the review might lead to a change in rating. Any changes have to be approved by the rating committee before it takes effect. The S&P's rating process is illustrated in Figure 2.

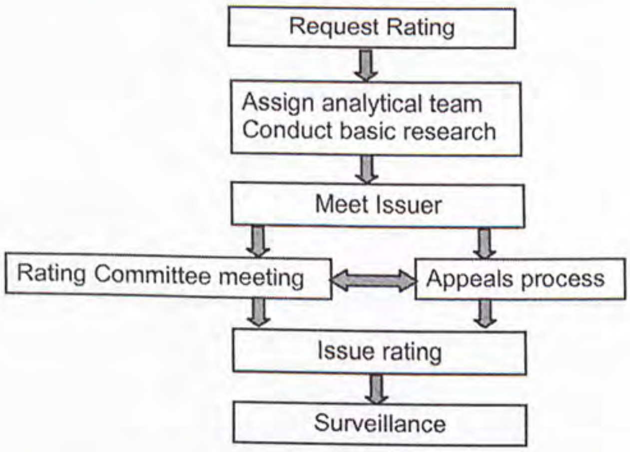


Figure 2: S&P's rating process (S&P, 2002)

1.2 Research Approach and Design

The main research issue is how to combine due diligence treatment into credit risk evaluation, and develop a reliable credit rating system for firms in China. A study framework – COMPLETE is proposed as a guideline for developing such a system. The credit rating consideration is decomposed into 8 aspects, and two practical due diligence treatments were proposed: in the first design, financial data reliability is evaluated as a *trustworthiness score*, and the bankruptcy prediction model developed with this score would be robust to manipulated data; in the second design, financial data reliability assessment is an independent subsystem, which is able to provide expertise more flexibility and decision supports.

The rest of this research is focused on quantitative analysis, which is aimed at providing helpful solutions and findings to credit rating system implementation. Firstly, we used Univaiate analysis to look for key indicators that discriminate bankruptcy from non-bankruptcy Chinese firms. Usefulness of 17 financial ratios previously applied to evaluate Chinese firms' performance was reexamined in this study. Logit regression is applied to develop a bankruptcy prediction model from 72 A-share companies listed on Shanghai and Shenzhen exchange covering period 1998 - 2003. The bankruptcy prediction model is then applied on one of the largest financial scandal cases in China, and a large discrepancy is observed from the credit score obtained from true and falsified data. We illustrated how to apply the COMPLETE framework and design analytical tests to discover the financial statement frauds. We further applied Classification and Regression Tree (CART) to analysis 22 false financial reports (FFS), and some interesting findings were found quite useful for FFS modeling and detection.

1.3 Organization of the Thesis

Organization of the rest of this thesis is as follows: Chapter 2 reviewed literatures in credit risk modeling; Chapter 3 explored empirical studies in financial scandal and formal approaches in FFS detection; Chapter 4 discussed credit scoring and FFS detection methodologies in our research; Chapter 5 described the COMPLETE framework and two practical designs proposed to deal with due diligence problems; Chapter 6 presented the development of the credit scoring model for firms in China and discussed observations from our statistical analysis; Chapter 7 discussed a number of FFS investigation efforts and highlighted several interesting findings; and Chapter 8 summarized this thesis.

CHAPTER 2
LITERATURE REVIEW ON CREDIT RISK MODELING

2.1 Overview

Credit risk evaluation is one of the hottest research areas in financial risk management, numerous studies have been conducted (Beaver 1966, Altman 1968, Deakin 1972, Edmister 1972, Altman et al. 1977, Ohlson 1980, Morris 1983, Zimmermann 1987, Tam & Kiang 1992, Quinlan 1993, Altman et al. 1994). Generally speaking, there are two schools of methodologies: Credit Value-at-Risk (CVaR) and credit scoring.

In CVaR methodologies, probability of default (PD) is derived from credit risk models built upon the underlying dynamics of interest rates and firm characteristics. Approaches in this category include **the Credit Migration Approach, the Actuarial Approach, Discrete Time Multi-period Model, and Structural Approach:**

1. **Credit Migration Approach** is based on the probability of moving from one rating category to another with the transition modeled as a stationary Markov-process; it is proposed by JP Morgan and implemented in CreditMetrics (CreditMetrics 1997).
2. **The Actuarial Approach** assumes the default for loans follow an exogenous Poisson process, and an application is CreditRisk+ (Credit Suisse 1997).
3. **The Discrete Time Multi-period Model** is built upon the assumption that default probabilities are conditional on macro-variables; this approach is proposed by McKinsey and implemented in its product CreditPortfolioView.
4. **The Structural Approach** is based on the *option pricing model* proposed by Merton (Merton 1974). In this approach, the firm's equity is viewed as a call option on the firm's assets, with strike price equals to the book value of the its liabilities. The assets' fluctuation is assumed to follow a geometric Brownian motion, and the probability of default equals to the probability of the assets' value falls below some critical level. KMV (Moody's 1993) used this approach to calculate the Distance-to-Default (DD) and Expected-Default-Frequency (EDF).

Although the CVaR methodology is well developed and commercially successful, a common drawback is that firm-level information utilized is too little; for example, KMV model only considered a firm's market value of assets and book value of liabilities. On the contrary, credit scoring methodology considered an entire profile of characteristics of the relevant firms to predict bankruptcy rates and estimate the severity of losses. Credit scoring methodology can be further classified into three broad categories: the first category is **Multivariate Statistical Approach** which tries to derive a linear or non-linear combination of several independent variables to best-discriminate between bankruptcy firms and non-bankruptcy firms, a large number of researches had been conducted with this approach (Beaver 1966, Altman 1968, Deakin 1972, Edminster 1972, Altman et al. 1977, Ohlson 1980, Morris 1983); the second category is **Expert System Approach**, which is aimed at constructing predictive models based on rules from a "knowledge-base" (Zimmermann 1987). However, this approach is not popular because the knowledge-base is hard to derive in practice; the third category is **Neural Networks (NN) Approach** which is based on more complicated mappings between the predictive and response variables (Quinlan 1993, Altman et al. 1994). It is one of the hottest research areas in the 90s and lots of debates had been made. Pros declared that NN produced better results for many classification and prediction problems empirically, and Cons argue that these is lack of logic or rule-based explanations for the input-output relationships, so NN is difficult to explain the underlying decision principle for rejecting applications (West 2000, Tam & Kiang 1992).

With above discussions, the multivariate statistical approach is considered most suitable for this research purpose: firstly, it included an entire profile of firm characteristics, and it is possible to dig out important indicators to predict bankruptcy for firms in China; secondly, the relationships between the input-output variables can be explained by solid statistical Theorems, and the results could help us find interesting phenomenon associated with Chinese firms. Therefore, the rest of the reviews are focused on studied using three multivariate statistical approaches, namely discriminate analysis (DA), Logit regression and classification and regression tree (CART). When applying these techniques, two concerns are raised: how to select variables and how to test the model performance. These topics are also discussed in this Chapter.

2.2 Discriminant Analysis

Discriminant analysis (DA) is a statistical technique used to classify an observation into one of several pre-defined groupings depending upon the observation's individual characteristics. The simplest form of DA is linear discriminant analysis (LDA), which attempts to derive a linear combination of these characteristics. The LDA function is:

$$Z = V_1X_1 + V_2X_2 + \ldots + V_nX_n$$

where V_1, V_2, ... V_n are discriminant coefficients and X_1, X_2, ... X_n are independent variables. The model is typically fit via least squares with the objective of maximizing inter-group to intra-group variance.

In 1968, Altman built the well-known Z-score model with LDA. His sample was consisted of 66 manufacturing firms dividing into two groups: the bankruptcy group had 33 firms filed under the National Bankruptcy Act from 1946 to 1965, one-year before bankruptcy data are used. The non-bankruptcy group was consisted of 33 matched firms, which were chosen on a stratified random basis – stratified by industry, asset size and year. 22 potential helpful financial ratios were selected as the initial profile of variables. The final profile was selected through four steps: (1) observe statistical significance of each independent variable; (2) evaluate correlations among the relevant variables; (3) observe the predictive accuracy of the various profiles; and (4) judgment of the analysts. The final discriminant function took the form of:

$$Z = 1.2X_1 + 1.4X_2 + 3.3\,X_3 + 0.6X_4 + 1.0X_5$$

where:

X_1 = Working capital / Total assets

X_2 = Retained earnings / Total assets

X_3 = Earnings before interest and taxes (EBIT) / Total assets

X_4 = Market value equity / Book value of total debts

X_5 = Sales / Total assets

The best critical value falls between 2.67-2.68 and 2.675 was chosen as the cut-off score. The model achieved 95% overall prediction accuracy tested from hold-out data.

In 1972, Deakin developed an alternative business failure prediction model with LDA. The failure group in his analysis contained 32 firms failed between 1964 and 1970, and he employed similar paired sample strategy to form the non-failure group. He applied 14 financial ratios formerly used by Beaver (Beaver 1968) to develop the discriminant model. His study found that discriminant analysis could make accurate predictions within 3-year horizon. His model applied on 1-year, 2-year and 3-year before failure data achieved 96%, 95.5%, and 95.5% overall prediction accuracy respectively.

Also in 1972, Edmister attempted to apply LDA to predict failure of small business firms. In his study, failure was defined as the recognition of a loan loss. 84 samples were drawn from the Small Business Administration (SBA) loan recipients covering the period 1954-1969, where 42 observations were loss cases and the other half were non-loss cases. The pool of variables was consisted of 19 financial ratios, and trends and industry positions were also considered. A stepwise LDA was applied and the following model was obtained:

$$Z = 0.951 - 0.423X_1 - 0.293X_2 - 0.482X_3 + 0.277X_4 - 0.452X_5 - 0.352X_6 - 0.924X_7$$

where:

X1=1 if funds flow / current liabilities <0.05; 0 otherwise

X2=1 if equity / sales <0.07; 0 otherwise

X3=1 if (net working capital / sales) / industry average < - 0.02; 0 otherwise

X4=1 if (current liabilities / equity) / industry average < .48; 0 otherwise

X5=1 if (inventory / sales) / industry average < .04 and trends upward; 0 otherwise

X6=1 if quick ratio / industry average < .34 and trends downward; 0 otherwise

X7=1 if quick ratio / industry average trends upward; 0 otherwise

The model achieved an overall classification accuracy of 93% and it appeared to be successful. This study also illustrated that the predictive power of ratio analysis depends upon both the choice of analytical methods and selection of ratios: there was no single ratio which predicts nearly as well as a small group of ratios; furthermore, some ratios that appear not significant in the univariate analysis may serve to improve overall prediction accuracy when added to the discriminant function.

Although LDA appeared to be a successful method in bankruptcy studies, several concerns had been raised: firstly, the independent variables are assumed to follow multivariate normal distribution, however, the real data are not distributed normally (Eisenbeis 1977, McLeay 1986); secondly, the linear classification rule is only optimal when two groups has equal covariance matrices, however, this condition is not usually hold (Hamer 1983); Moreover, the output of discriminant analysis is basically an ordinal ranking/discriminatory device, which has little intuitive interpretation (Ohlson 1980).

2.3 Logit Regression

If LDA was the dominant approach in bankruptcy studies during the 60s and 70s, then Logit regression replaced its role during the 80s and 90s. In 1980, Ohlson first applied Logit regression to bankruptcy studies. His sample contained 105 bankruptcy firms and 2058 non-bankruptcy firms covering the period 1970-1975. Data for these firms were derived from the 10-K financial statements. Two Logit models were developed to predict probability of bankruptcy 1-year later and 2-year later; the prediction accuracy was 96.12% and 95.55% respectively. His study identified four statistically significant factors in affecting the probability of failure: the size of the company, measures of the financial structure (total liabilities to total assets, working capital to total assets), a measure of performance (net income to total assets) and a measure of current liquidity (funds from operations to total liabilities). He also argued that the matching principle used in previous studies did not reveal any superiority to non-matching strategy, which is solely based on large number Theorem.

There are two major differences between Logit regression and LDA: firstly, there is no assumption on multivariate normal distribution and equal covariance matrices; secondly, Logit regression is fit via maximum likelihood learning, and the output is a probabilistic measurement rather than a simple ranking or score. The regression coefficients also have interpretations associated with this probability (details of Logit regression is discussed in Chapter 5). Some researches indicated that Logit regression produced more promising results (Lennox 1999, West 2000). However, Logit regression is not as commercially successful as Altman's Z-score.

2.4 Regression and Classification Tree (CART)

Regression and Classification Tree (CART) is a tree-building technique which is unlike traditional data analysis methods (details of CART is discussed in Chapter 6). In 1985, Frydman et al. first applied CART to bankruptcy studies. Their sample was consisted of 58 bankrupt industrial firms and 142 randomly selected non-bankrupt manufacturing and retailing firms covering the period 1971-1981. Data were obtained from COMPUSTAT. 20 financial ratios were evaluated and the following tree was obtained:

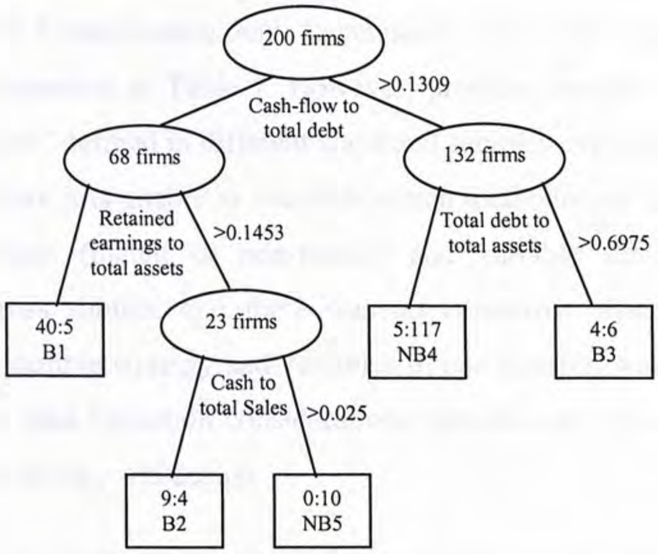


Figure 3: CART based on financial data on 200 firms (Frydman et al 1985)

Prior probabilities of bankrupt (group 1) and non-bankrupt (group 2) groups were specified as 0.02 and 0.98 respectively, and misclassification costs were specified as $c_{12} = 50$, $c_{21} = 1$. The terminal node corresponded to group 1 was denoted by B (bankrupt) and group 2 was denoted by NB (non-bankrupt). 45 firms had fallen into the leftmost terminal node, in which 40 belong to group 1 and 5 belong to group 2. This study indicated that while both CART and the LDA lead to rather accurate classifications, the CART usually dominated LDA. The magnitude of dominance is dependent on the specification of costs of misclassification and prior probabilities.

CART is inherently non-parametric, which is a very desirable feature: no assumptions are made on the underlying distribution of the independent variables, and highly skewed or multi-modal data can be handled. Besides, the output of CART is intuitive for non-statisticians to interpret. However, this technique may suffer from over fitting because the output only carries binary value and lack of a probability estimate for the classification. Moreover, CART was not as popular as LDA and Logit regression in bankruptcy studies.

2.3 Chapter Summary

This Chapter reviewed methodologies applied in bankruptcy prediction and credit risk assessments. Two schools of methodologies – credit Value-at-Risk (CVaR) and credit scoring were introduced; three statistical approaches – discriminant analysis, Logit regression and Classification and Regression Tree were discussed in detail. A summary is presented in Table 1. However, previous studies were conducted with "business failure" defined in different ways and samples extracted from different time periods, therefore it is unable to conclude which methodology is superior. Moreover, sample strategies (match or non-match) and variable selections were varying arbitrarily across studies, and there was no consensus. Therefore, the choice of methodology, sample strategy and variables in our research was stick to the research objectives and data limitation considerations. Details were discussed in Chapter 5 – research methodology and design.

Model	LDA	Logit Regression	CART
Assumption	• Multivariate normal distribution and equal covariance matrices	• Cumulative logistic distribution	• None
Output	• Discrimant score/ Ranking	• Probability of bankruptcy	• Class/Category
Previous Researches Reviewed	• Altman 1968 • Deakin 1972 • Edmister 1972	• Olhson 1980	• Frydman et al 1985
Commercial Applications	• Z-score	• N/A	• N/A
Pros	• Good classification capability • Popular	• Good classification and prediction capability • No constraints on distribution of independent variables • Good interpretation ability • Popular	• No constraints on distribution of independent variables • Result intuitive to interpret
Cons	• Underlying assumption may not be satisfied	• N/A	• Over fitting • Not popular

Table 1: Summary of Models

CHAPTER 3

FALSE FINANCIAL STATEMENTS DETECTION

3.1 Overview

It is not exaggerated to criticize that no one can rely on Chinese firms' financial statements because the books are cooked. From March 1999 to December 2001, the Ministry of Finance (MOF) organized several investigations and the results were astonishing: according to inspection bulleting No. 5 (10-July-2001), 147 out of 159 investigated firms provided untrue asset values, 155 provided untrue equity values, and 157 provided untrue incomes; the proportion of falsified amount comprised 0.95%, 1.82% and 33.4% with respect to corresponding true values. Inspection bulleting No. 7 (12-December-2001) revealed that 160 out of 320 investigated organizations reported more than 1% falsified assets and 183 reported more than 10% falsified incomes; total overstated amount for assets was RMB 7.38 billion, and that for incomes was RMB 3.51 billion. In the same year, the National Auditing Office (CNAO) inspected 32 auditing reports provided by 16 auditing firms, it was found that 14 companies (including 41 chartered accountants) provided 23 auditing reports with serious inconsistent opinions, and the total amount of fabricated figure added up to RMB 7.1 billion. However, it might be just "the tip of the iceberg": according to MOF's inspection bulleting No. 5, there were no more than 100 accounting scandal cases discovered by the China Securities Regulation Commission (CSRC) during the last 10 years, however, the total number might be over 10,000, implies that the probability of "caught and punishment" was at most 1%.

Manipulating financial statements had been an extremely serious problem affecting investors' confidence towards firms in China. Some people criticize that the problem is due to loose supervision and a lenient punishment mechanism towards Chinese accounting professionals (Lipsher 2002), other people argue that the situation is caused by inadequate corporate governance and inappropriate ownership-responsibility structure that creates insider-control opportunities (Qiu Yigan 2003). The survey of Hong Kong Corporate Governance Council in January 2003 summarized seven major problems affecting credibility of Chinese firms: (1) insider control and concentration of power; (2) low business ethical standard and lack of proper corporate cultures; (3) lack of independent board and effective controls; (4)

lack of incentives and mature labor market for selecting executives; (5) weak regulatory enforcement; (6) low corporate transparency and disclosure quality and (7) shortage of independent and quality auditors and other intermediaries

Therefore, financial data quality should be carefully examined before applied for credit risk assessments, and in order to do that, techniques associated with financial data manipulation and characteristics with false financial reports should be understood. The next Section reviewed empirical studies on financial scandal and financial statement fraud.

3.2 Empirical Studies on Financial Scandal and False Financial Statements (FFS)

Financial scandal and false financial statements had long been concerned worldwide. The techniques associated with producing FFS were discussed in Schilit's book "Financial Shenanigans" (Schilit 1993), and seven common tricks were: (1) recording revenue before it is earned; (2) creating fictitious revenue; (3) boosting profits with non recurring transactions; (4) shifting current expenses to a later period; (5) failing to record or disclose liabilities; (6) shifting current income to a later period and (7) shifting future expenses to an earlier period. The first five tricks could boost current year earnings, and the last two were balancing acts in order to make financial accounts grow smoothly. Apart from them, following tricks made detection more difficult for firms in China (Feicao 2001):

1. **Related party transactions:** companies may move large amount of deficit into their related parties. Qiong Minyuan – one of the largest financial scandal cases in China, used to employ this technique and boost 540 million incomes.
2. **Non-monetary and special transactions:** non-monetary transactions (such as transferring land ownership and stock ownership) do not create cash-flow, but the intangible gains are recognized in income. Profits obtained from these operations are highly sensitive to manipulations.
3. **Assets restructuring:** assets restructuring has many positive impacts for company expansion, however, some Chinese firms make it into *financial statement restructuring* or *performance enhancement*. The accounting process involved is very complicated, and difficult for auditors to detect

4. **Change of accounting estimates:** the Chinese accounting principle allows companies to change their accounting estimates if necessary. Some firms take it as a mean to manipulate income. Typically, it is achieved by changing the method for estimating long-term investments and adjusting the area of consolidation.

In order to detect FFS, a number of fundamental studies had been conducted. A report by the Committee of Sponsoring Organizations of the Treadway Commission examined fraudulent financial reporting from US public companies covering the period 1987-1997, and some critical insights include (Beasley 1999):

1. The companies committing fraud generally were small
2. In 72% of the cases CEO appeared to be associated with the fraud, and in 43% of the cases CEO were associated with financial statement fraud
3. The audit committee was weak, and the board was dominated by insiders
4. The founders and board members owned a significant portion of the company
5. Sever consequences were expected when frauds were committed, including bankruptcy, significant changes in ownership, and/or other serious punishments

In 1997, the American Institute of Certified Public Accounts (AICPA 1997) issued Statement of Auditing Standards (SAS) No. 82: *Consideration of Fraud in a Financial Statement Audit*. This statement discussed risk factors, or *red flags* related to fraudulent financial reporting. The signals were grouped into three categories, **management's characteristics and influence over the control environment**, **industry conditions**, and **operating characteristics and financial stability**. Risk factors associated with **Management's characteristics and influence over the control environment** included:

1. A significant portion of management's compensation is represented by bonuses, which is contingent upon the entity's operating results or financial position.
2. An excessive interest by management in maintaining or increasing the entity's stock price or earning trend through aggressive accounting practices.
3. An interest by management in pursuing inappropriate means to minimize reported earnings for tax-motivated reasons.
4. Domination of management by a single person or small group without compensating controls such as effective oversight by the board of directors or audit committee.

5. Inadequate monitoring on significant controls.
6. Management sets aggressive financial target and expectations for operating personnel.
7. Management continues to employ an ineffective accounting, information technology or internal auditing staff
8. High turnover of senior management, counsel, or board members
9. Known history of securities law violations
10. Switching CPA and/or conducting asset re-evaluation for unknown reasons.

Risk factors associated with **Industry conditions** included:

1. New accounting or regulatory requirements that could impair the financial stability or profitability of the entity.
2. High degree of competition or market saturation, accompanied by declining margins
3. Declining industry with increasing business failures and significant declines in customer demands.
4. Rapid changes in the industry, such as high vulnerability to technology advancements and/or fast product obsolescence.

Risk factors associated with **Operating characteristics and financial stability** included:

1. Significant pressure to obtain additional capital necessary to stay competitive considering the financial position.
2. Assets, liabilities, revenues, or expenses based on significant estimates that involve unusually subjective judgments or uncertainties, or that are subject to potential significant change in the near term in a manner that may have a financially disruptive effect on the entity, such as ultimate collectibility of receivables, timing of revenue recognition, realizability of financial instruments based on the highly subjective valuation of collateral, difficult-to-assess repayment sources or significant deferral of costs.

Similar study had been conducted in China. An article (Feicao 2001) indicated that four types of companies are most prone to managerial scandals:

1. **Companies with frequent capital operations and related-party transactions:** most of these operations were fabricated.
2. **Companies with high and volatile stock prices:** the illusion was typically created by market makers and the performance was just a lie.
3. **Initial Public Offering (IPO) companies:** they were usually restructured before listing, and performance reported in the prospectus was typically far too good from their true capabilities.
4. **Companies in a declining or over competitive business environment:** these companies usually had poor actual performance, however, they employed accounting tricks to create stable income over a few years, and then they simply announced a big drop in performance and quit the market with all the capital raised.

FFS could seriously affect accuracy of the bankruptcy prediction models, unless the data are cleaned and processed at the surface of truth, the results cannot be reliable. However, knowing what item and how much is it manipulated is almost impossible unless legal entities investigated into the company's accounts, therefore a challenge is raised on how to evaluate financial statements reliability with only published data. The rest of this Chapter reviewed researches related to this issue.

3.3 False Financial Statements (FFS) Detection

Recent studies attempted to build FFS detection models and three methodologies: Logit regression, artificial neural network (ANN) and fuzzy neural network (FNN) were employed.

3.3.1 Logit Regression

Beasley (Beasley 1996) used Logit regression to test the prediction that the inclusion of larger proportions of outside members on the board of directors significantly reduces the likelihood of financial statement fraud. 75 FFS samples were drawn from the Accounting and Auditing Enforcement Releases (AAER) issued by the SEC covering the period 1980-1991, and the non-FFS group is composed of 75 matched samples by industry, financial size, year and stock exchange listed. 12 variables were examined:

1. %OUTSIDE = the percentage of outside directors on the board
2. OUTOWNBD = the percentage of outstanding common stock shares held by outside directors
3. OUTTENURE = the mean number of years that outside directors served on the board
4. DIRECTSHIP = the mean number of additional directorships held by outside directors
5. BOARDSZ = the number of members on the board
6. GROWTH = the average percentage change in total assets for two years ending before the year of financial statement fraud
7. TROUBLE = 1 if the firm reported at least three annual net losses in the six-year period preceding the first year of financial statement fraud, 0 otherwise
8. AGEPUB = the number of years the firm's stock had traded
9. MGTOWNBD = the cumulative percentage of ownership in the firm held by insiders who serve on the board
10. CEOTENURE = the number of years that the CEO has served as a director
11. BOSS = 1 if the chairperson of the board is the CEO or president, 0 otherwise
12. BLOCKHLD = the cumulative percentage of outstanding common stock shares held by block-holders (holding > 5%) who are not affiliated with management

Results from Logit regression analysis indicated that non-fraud firms have boards with significantly higher percentages of outside members than fraud firms. In addition, as outside director ownership and outside director tenure, and/or number of outside directorships held by outside directors decreased, the likelihood of financial statement fraud decreased. Other variables were not statistically significant.

Spanthis (Spathis 2002) employed stepwise logistic regression to construct FFS detection models for firms in Greece. His sample was consisted of 38 false reports and 38 matched non-false reports from the manufacturing sector. Ten financial ratios were included: (1) debt to equity, (2) sales to total assets, (3) net profit to sales, (4) accounts receivable to sales, (5) net profit to total assets, (6) working capital to total assets, (7) gross profit to total assets, (8) inventory to sales, (9) total debt to total assets, and (10) financial distress (Z-score). Five variables, namely (1) debt to equity, (2) sales to total assets, (8) inventory to sales, (9) total debt to total assets, and (10) Z-score were found helpful to discriminate false and non-false financial statements. Fraudulent financial statements were associated with higher debt to equity, lower sales to total assets, higher inventory to sales, higher total debt to total assets, and lower Z-score comparing to non-fraud ones. Two models were constructed from stepwise logistic regression, the first model consisted of (8) inventory to sales, (5) net profit to total assets and (6) working capital to total assets was able to achieved an overall prediction accuracy of 82.89%; the second model consisted of (8) inventory to sales, (9) total debts to total assets and (10) Z-score achieved a higher overall prediction accuracy (84.21%). Both models were statistically significant.

3.3.2 Artificial Neural Networks (ANN)

Artificial Neural Networks (ANNs) are computer programs that simulate the way human brain functions. Analogous to human brains, ANNs are consisted of simulated neurons in the format of layers and nodes. The main advantage of ANN is its ability to learn through a trial-and-error process, in which neurons adjust their weights for input variables and model the behavior or patterns of output variables.

ANN was first employed to predict management fraud by Fanning and Cogger. (Fanning & Cogger 1998), their study examined 62 variables that cover **(1) corporate governance, (2) auditor, agency issues, (3) subsidiaries, (4) capital structure, (5) operating results, (6) personnel, (7) litigation, (8) accounting choices, (9) financial ratios, and (10) financial accounts and trends:**

1. Corporate governance

1.1. Number of members on the board of directors
1.2. Percentage of outside directors on the board
1.3. If the board had a director who was an academia
1.4. If the board had a director who was an lawyer
1.5. If the CEO was also the chairman
1.6. If the company had an audit committee
1.7. If the company had an executive committee
1.8. If the company had a compensation committee
1.9. If the company had a nomination committee

1.10. Percentage of outsiders on the audit committee

1.11. Percentage of outsiders on the executive committee

1.12. Percentage of outsiders on the compensation committee

1.13. Percentage of outsiders on the nomination committee

2. Auditor

2.1. If the company changed auditor from last year

2.2. If the company changed auditor over the last 3 years

2.3. If the company had qualified the audit report in the current year

2.4. If the company had a qualified audit report in the past 3 years

2.5. If the company had a non-Big Six auditor

3. Agency Issues

3.1. Percentage of stock ownership by officers and directors

3.2. Percentage of stock ownership by CEO

3.3. If the company had a short-term compensation plan

3.4. If the company had a long-term compensation plan

3.5. If the company had a profit sharing plan

4. Subsidiaries

4.1. Number of subsidiaries

4.2. If the company had two or more foreign subsidiaries

5. Capital Structure & Operating results

5.1. Z-score (Altman)

5.2. Geometric growth rate of sales for the last 3 years

6. Personnel

6.1. If the company changed CFO from last year

6.2. If the company changed CFO over last 3 years

6.3. If the company changed CEO from last year

6.4. If the company changed CEO over last 3 years

6.5. If the CEO and CFO was the same person

6.6. If the company had a family relationship on the board or executive

7. Litigation

7.1. If the company had litigation against it over the last 3 years

8. Accounting Choices

8.1. If the company used the LIFO method for inventory

8.2. If the company used an accelerated depreciation method

9. Financial Ratios

9.1. Accounts receivable to sales

9.2. Accounts receivable to total assets

9.3. Gross margin = 1 – cost of goods sold / sales

9.4. Inventory to total assets

9.5. Inventory to sales

9.6. Allowance for doubtful accounts to accounts receivable

9.7. Allowance for doubtful accounts to sales

9.8. Net property, plant, and equipment to total assets

9.9. Debt to equity

9.10. Net income before tax to sales

9.11. Net income before tax to stockholders' equity

9.12. Sales to total assets

10. Financial Accounts & Trends

10.1. Accounts receivable

10.2. Allowance for doubtful accounts

10.3. Inventory

10.4. Net property, plant and equipment

10.5. Log(total assets)

10.6. Log(total debt)

10.7. Retained earning

10.8. Working capital

10.9. Sales

10.10. Cost of goods sold

10.11. Earnings before interest and taxes

10.12. Extraordinary items

10.13. If company's 1-year account receivable growth exceed 10%

10.14. If the company's 1-year gross margin growth excess 10%

The analysis trough AutoNet (one of ANN implementations) using 102 FFS and 102 matched non-FFS from SEC found 20 possible FFS indicators, they were (1.1) board size, (1.2) percentage of outside directors, (1.5) chairperson also the CEO, (1.6) existence of an audit committee, (1.8) existence of a compensation committee, (2.5) having a non-Big Six auditor, (3.5) having a profit-sharing plan, (5.1) financial distress (Z-score), (5.2) growth rate, (6.2) change in chief financial officer in the last three years, (6.5) having the president as the treasurer, (7.1) involvement in litigation, (8.1) choice of the LIFO inventory method, (9.1) the ratios of accounts receivable to sales, (9.5) inventory to sales, (9.8) net property plant and equipment to total assets, (9.9) debt to equity, (9.12) sales to total asset and growth rate larger than 10% for (10.13) accounts receivable and (10.14) gross margin. It is found that FFS were associated with higher accounts receivable to sales, higher inventory to sales, lower net property plant and equipment to total assets, higher debt to equity, lower Z-score, lower sales to total assets, and higher accounts receivable growth rate. The analysis also indicated that ANN out performed traditional statistical techniques such as LDA and Logit regression in total predicting accuracy.

3.3.3 Fuzzy Neural Networks (FNN)

Fuzzy Neural Networks (FNNs) are a class of hybrid intelligent systems that integrate *fuzzy logic* with ANNs. Fuzzy logic is a logical system used to operate on *fuzzy sets*, which were developed to represent, manipulate and utilize uncertain information and provide a framework for handling uncertainty and imprecision in real-world applications. For example, conventional rule-based systems usually fix 10% as a threshold to detect anomalous account trends, but a fuzzy logic system is able to dynamically adjust the threshold by studying the entire range of possible deviations from the training datasets.

Jerry et al. (Jerry et al. 2003) developed a FFS detection model based on fuzzy clustering (GENFIS2 function of the Matlab Fuzzy Logic Toolbox), and the model was further tuned by Adaptive Neuro-Fuzzy Inference System (ANFIS function of the Matlab Fuzzy Logic Toolbox). The sample was consisted of 40 FFS and 160 non-FFS (1 FFS is matched to 4 non-FFSs) matched against industry, firm size and year. The sample covered the period 1980-1995. Variables include: (1) allowance for doubtful accounts to net sales; (2) allowance for doubtful accounts to accounts receivable; (3) accounts receivable to net sales; (4) accounts receivable to total assets; (5) gross margin to net sales; (6) net sales; (7) accounts receivable; and (8) allowance for doubtful accounts. Data were extracted from Standard & Poor's Research Insight database. The testing results indicated the FNN achieved significantly higher accuracy in predicting fraud cases than Logit regression (35% v.s. 5%).

3.4 Chapter Summary

This Chapter reviewed empirical and formal studies in false financial statements (FFS) detection. These studies suggested a wide range of possible FFS indicators, which are summarized in Table 2.

Category	Indicators	Remarks
Financial Ratios & Account Trends	• Accounts receivable / revenue • Gross margin / revenue • Account receivable / total assets • Cost of goods sold / revenue • Revenue / total assets • Revenue / inventory • Net income / net worth • Net income / total assets • Net income / revenue • Cash from operations / revenue • Cash from operations / accounts receivable • Cash from operations / income from operations • Cash from operations / inventory • Revenue • Net income • Accounts receivable • Inventory	• Accounts affected by transactions in the revenue cycle were prone to FFS, and related accounts (such as revenue, account receivables, inventory, income, etc) and financial ratios should be closely examined. • The trend should be benchmarked to industry average and similar situated firms, any anomalies found should be suspected.
Corporate Governance	• Percentage of outsiders on the board • Presence of lawyers or academics on the board • Existence and structure of the committees	• Outside directors may provide better monitoring • Having a lawyer or academics on the board may mitigates the chances of corporate malfeasance

	• Whether CEO and chairman is the same person	• Inclusion of audit committee, nominating committee and compensation committee may mitigate FFS
Auditor	• Auditor quality • Auditor change • Audit's opinion	• Large audit firms are considered to be more qualified than small ones • Auditor change may be close associated with FFS
Agency Issues	• Percentage of stock ownership by top management • CEO stock ownership • Short-term bonus plans	• High concentration of ownership is an indicator of insider-control • Existence of short-term bonus plan may trigger management fraud
Subsidiaries	• Number of subsidiaries • Existence of foreign subsidiaries	• Companies with large number of subsidiaries or foreign subsidiaries are prone to FFS
Personnel	• CEO turnover • CFO turnover • Whether CEO is also the treasurer • Whether family relationships presents in directors and officers	• Higher turnover of CEO and CFO maybe a signal of FFS • Concentration of duty may cause FFS
Litigation	• Existence of litigation	• The existence of litigation during past years maybe a warning signal for FFS

Table 2: Summary of FFS Indicators

Similar to credit scoring, FFS detection and prediction belongs to the category of classification and decision problems, so traditional statistical techniques and more advanced ones such as Artificial Neural Networks (ANNs) and Fuzzy Neural Networks (FNNs) can all be applied. Empirical studies indicated that ANNs and FNNs produced better prediction accuracy than statistical approaches, however, this conclusion should be used cautiously because ANNs and FNNs algorithms already combined optimization procedures into the model construction process, therefore, it is more adaptive to the datasets by nature. Moreover, manipulated financial statements exhibited different characteristics due to variations in business fields and regulatory requirements on accounting methods and disclosure requirements. Therefore, a comprehensive investigation about FFS in Chinese firm is necessary. Some discussions were presented in Chapter 5 and Chapter 7.

CHAPTER 4

RESEARCH METHODOLOGIES IN CREDIT SCORING & FALSE FINANCIAL STATEMENTS DETECTION

4.1 Overview

Multivariate statistical techniques categorize credit scoring problems to a broader category called *classification and decision problems*. This kind of problem is consisted of four major components: (1) the categorical *outcome* or *dependent variable*, which typically takes binary value, such as "bankrupt" and "non-bankrupt"; (2) *independent variables*, which are characteristics that potentially related to the dependent variable. For bankruptcy prediction, independent variables could be a number of firm-level and/or macroeconomic variables; (3) the learning dataset used for model development, and (4) the testing dataset used to verify the model. Formal studies in FFS detection were aimed at finding a sound discriminate scheme along with the discriminators that best separate false statements from non-false ones, therefore, methodologies applied in credit scoring are also applicable to FFS detection.

Before applying the multivariate techniques, statistical significance of independent variables are typically examined through univariate tests, where the *t*-statistics is employed to assess whether the means of two groups are *statistically different* from each other. The formula of *t*-statistics is:

$$t - value = \frac{mean\,(group\,1) - mean\,(group\,2)}{\sqrt{\dfrac{\mathrm{var}\,iance\,(group\,1) + \mathrm{var}\,iance\,(group\,2)}{n-1}}}$$

where n is the number of observations in each group (assume that the size of two groups are equal). The numerator of the ratio is just the difference between the two means, and the denominator is the *standard error of the difference* – a measure of the variability or dispersion. Larger *t*-value implies that larger difference between two means and/or smaller standard error of the difference (smaller spread). If the *t*-value is larger than the critical value specified by some confidence level (such as 95%) in the *t*-table, then the corresponding indicator is concluded as *statistically significant* in discriminating two groups.

4.2 Logit Regression

Over the last two decades, the Logit regression model had become the standard method of analysis concerned with describing the relationship between a response variable and one or more explanatory variables. Although it sounds similar to linear regression, the difference is reflected in both the choice of the *parametric model* and the *underlying assumptions*. Typically, Logit analysis is target to *binary* or *dichotomous* outcome variable, and the parameters are estimated through *maximum likelihood estimation*. In this Section, we first discuss how Logit regression estimate its parameters for the single explanatory variable case, and then discuss the generalized Logit model and how to apply and evaluate it in bankruptcy studies.

The Logit (logistic) distribution function takes the form of: $F(x) = (1 + \exp(-x))^{-1}$ Let $(Y_1, X_1),\ldots,(Y_n, X_n)$ be a random sample from the conditional Logit distribution:

$$P[Y_j = 1 \mid X_j] = (1 + \exp(-\alpha_0 - \beta_0 X_j))^{-1}$$
$$P[Y_j = 0 \mid X_j] = 1 - (1 + \exp(-\alpha_0 - \beta_0 X_j))^{-1}$$

where α_0 and β_0 are unknown true parameters for the population. It is called the *Logit model*, because: $P[Y_j = 1 \mid X_j] = F(\alpha_0 + \beta_0 X_j)$

As the response variable of the Logit distribution is binary, we can imagine that the outcomes are randomly drawn from a *Bernoulli distribution*, which takes the form of:

$$P[Y_j = 1] = p_0 \text{ and } P[Y_j = 0] = 1 - p_0$$

where p_0 is assumed to be the unknown true underlying probability. The conditional probability function for Bernoulli distribution is:

$$\begin{aligned} f(y \mid p_0) &= P[Y_j = y \mid p_0] \\ &= p_0^{y} (1 - p_0)^{1-y} \\ &= p_0 \text{ if } (y = 1) \text{ and } (1 - p_0) \text{ if } (y = 0) \end{aligned}$$

Analogously, the conditional probability function for Logit distribution is:

$$f(y \mid X_j, \alpha_0, \beta_0) = P[Y_j = y \mid X_j, \alpha_0, \beta_0]$$
$$= F(\alpha_0 + \beta_0 X_j)^y (1 - F(\alpha_0 + \beta_0 X_j))^{1-y}$$
$$= F(\alpha_0 + \beta_0 X_j) \text{ if } y = 1 \text{ and } (1 - F(\alpha_0 + \beta_0 X_j)) \text{ if } y = 0$$

In order to estimate the parameters, *maximum likelihood estimation* is used, where the estimated parameters maximized the *likelihood function*. For Bernoulli distribution, the likelihood function is: $L_n(p) = f(Y_1|p).f(Y_2|p)\ldots f(Y_n|p)$, which is the joint probability that outcomes Y_1, Y_2, … Y_n are observed. The idea of maximum likelihood estimation is to choose a parameter p to maximize the joint probability, in other words the likelihood function $L_n(p)$ is maximized. Such p is an unbiased estimator of p_0. For convenience of the analysis, people usually consider the log-likelihood function, which is derived by taking nature log on both sides of the likelihood function. As log function is monotone increasing, so maximizing the log-likelihood function is equivalent to maximizing the original likelihood function. A more formal motivation for maximum likelihood estimation is based on the fact: $\ln(x) \le x - 1$ is valid for any positive x. Consequently, we have:

$$\ln\left(\frac{f(Y_j \mid p)}{f(Y_j \mid p_0)}\right) \le \frac{f(Y_j \mid p)}{f(Y_j \mid p_0)} - 1$$

holds for any Y_j. Take expectations on both sides, we have:

$$E\left[\ln\left(\frac{f(Y_j \mid p)}{f(Y_j \mid p_0)}\right)\right] \le E\left[\frac{f(Y_j \mid p)}{f(Y_j \mid p_0)}\right] - 1$$
$$= \frac{f(1 \mid p)}{f(1 \mid p_0)} P[Y_j = 1] + \frac{f(0 \mid p)}{f(0 \mid p_0)} P[Y_j = 0] - 1$$
$$= \frac{p}{p_0} p_0 + \frac{1-p}{1-p_0}(1-p_0) - 1$$
$$= 0$$

Therefore, we have:

$$E[\ln(f(Y_j \mid P))] - E[\ln(f(Y_j \mid P_0))] \le 0 \Rightarrow$$
$$E[\ln(L_n(p))] - E[\ln(L_n(p_0))] \le 0 \Rightarrow$$
$$E[\ln(L_n(p))] \le E[\ln(L_n(p_0))]$$

It is not difficult to observe that the expectation of the log-likelihood function $E[\ln(L_n(p))]$ is maximized when $p = p_0$. In other words, the parameters which maximized the log-likelihood function are regarded as good estimators of the true parameters. Back to the Logit model, the conditional log-likelihood function is:

$$\ln(L_n(\alpha,\beta)) = \sum_{j=1}^{n} \ln(f(Y_j \mid X_j,\alpha,\beta))$$

$$= \sum_{j=1}^{n} Y_j \ln(F(\alpha+\beta X_j)) + \sum_{j=1}^{n}(1-Y_j)\ln(1-F(\alpha+\beta X_j))$$

$$= -\sum_{j=1}^{n} Y_j \ln(1+\exp(-\alpha-\beta X_j)) - \sum_{j=1}^{n}(1-Y_j)\ln(1+\exp(-\alpha-\beta X_j)) + \sum_{j=1}^{n}(1-Y_j)\ln(\exp(-\alpha-\beta X_j))$$

$$= -\sum_{j=1}^{n}(1-Y_j)(\alpha+\beta X_j) - \sum_{j=1}^{n}\ln(1+\exp(-\alpha-\beta X_j))$$

where α and β are parameters to be estimated. Similarly, we have:

$$\begin{aligned}
E[\ln(\frac{f(Y_j \mid X_j,\alpha,\beta)}{f(Y_j \mid X_j,\alpha_0,\beta_0)}) \mid X_j)] &\le E[\frac{f(Y_j \mid X_j,\alpha,\beta)}{f(Y_j \mid X_j,\alpha_0,\beta_0)} \mid X_j)] - 1 \\
&= \frac{f(1 \mid X_j,\alpha,\beta)}{f(1 \mid X_j,\alpha_0,\beta_0)} P[Y_j = 1 \mid X_j] + \frac{f(0 \mid X_j,\alpha,\beta)}{f(0 \mid X_j,\alpha_0,\beta_0)} P[Y_j = 0 \mid X_j] - 1 \\
&= \frac{f(1 \mid X_j,\alpha,\beta)}{f(1 \mid X_j,\alpha_0,\beta_0)} f(1 \mid X_j,\alpha_0,\beta_0) + \frac{f(0 \mid X_j,\alpha,\beta)}{f(0 \mid X_j,\alpha_0,\beta_0)} f(0 \mid X_j,\alpha_0,\beta_0) - 1 \\
&= \frac{1}{1+\exp(-\alpha-\beta X_j)} + \frac{\exp(-\alpha-\beta X_j)}{1+\exp(-\alpha-\beta X_j)} - 1 \\
&= 0
\end{aligned}$$

In other words: $E[\ln(L_n(\alpha,\beta)) \mid X_j] \le E[\ln(L_n(\alpha_0,\beta_0)) \mid X_j]$ Therefore, a good approximation of the true parameters can be obtained by maximize the conditional log-likelihood function $[\ln(L_n(\alpha, \beta))]$. The first-order conditions for a maximum are:

$$0 = \frac{\partial \ln(L_n(\overline{\alpha},\overline{\beta}))}{\partial \overline{\alpha}} = -\sum_{j=1}^{n}(1-Y_j) + \sum_{j=1}^{n}\frac{\exp(-\overline{\alpha}-\overline{\beta}X_j)}{1+\exp(-\overline{\alpha}-\overline{\beta}X_j)}$$

$$0 = \frac{\partial \ln(L_n(\overline{\alpha},\overline{\beta}))}{\partial \overline{\beta}} = -\sum_{j=1}^{n}(1-Y_j)X_j + \sum_{j=1}^{n}\frac{\exp(-\overline{\alpha}-\overline{\beta}X_j)X_j}{1+\exp(-\overline{\alpha}-\overline{\beta}X_j)}$$

The maximum likelihood estimators can be solved from above equations numerically, which is available in most statistical computing software (such as SAS, SPSS, etc).

The general Logit model takes the form of:

$$P[Y_j = 1 \mid X_{1j}, X_{2j}, \ldots X_{kj}] = (1 + \exp(-\beta_0 - \beta_1 X_{1j} - \beta_2 X_{2j} \ldots - \beta_k X_{kj}))^{-1}$$

where Y_j is the binary outcome for observation j. β_1, β_2... β_k are regression coefficients and β_0 is the interception term. These parameters are obtained by maximum likelihood estimation as before.

When applying the Logit analysis to bankruptcy studies, the basic idea can be described as: given a firm belongs to some pre-specified population, what is the probability that the firm goes into bankruptcy within some pre-specified time period. Maximum likelihood estimation implies that the joint probability of outcomes is maximized via appropriate setting of the parameters (regression coefficients and intercept term), therefore, the coefficients obtained by Logit regression are more meaningful and insightful. One can observe that $P[Y_j = 1|X_j]$ is an increasing function of X_j. Suppose Y_j=1 stands for outcome j is bankruptcy, if some coefficient is positive, it implies the corresponding independent variable has an increased probability with bankruptcy, the larger the independent variable, the outcome is more likely to be bankruptcy; on the other hand, if the coefficient is negative, it implies a decreased probability with bankruptcy, the larger the independent variable, the outcome is more likely to be non-bankruptcy. This property is desirable for researchers looking for key indicators or driving forces in bankruptcy studies.

It is also important to test the significance of the Logit regression model developed. Typically, significance is tested against the null hypothesis that all parameters are equal to 0, in which the variance is solely caused by randomness and no explanatory variable is statistically significant. The guiding principle to conduct the significance test is: compare observed values of the response variable to predicted values obtained from models with and without the variable in question, see if the inclusion of the variable had a great improvement over accuracy of the model. In order to do it, *goodness-of-fit* should be defined first. In Logit regression, measurement of goodness-of-fit is based on the log-likelihood function: the likelihood of the current model is compared against a benchmark called the *saturated model*, which possessed following properties: (1) the saturated model used the same function as the model of interest (for example, both Logit); (2) the number of parameters in the saturated model is equal to the sample size. The comparison is based on the following expression:

$$D = -2\ \ln[\textit{likelihood of current model / likelihood of saturated model}]$$

The quantity inside the brackets is called the *likelihood ratio*. The reason to use above expression is to insure the result obtained has a Chi-squared distribution with N-k degrees of freedom, where N is the sample size and k is the number of variables (see Dobson 1990). The result D is called the *deviance*, which is a generalization of residual sum of squares. A small deviance implies the likelihood of current model is close to that of the saturated model, and the current model provided sufficient explanation over the variance; on the other hand, a large deviance would be observed. This statistics is provided by most statistical software (such as SAS, SPSS, Matlab Statistics toolbox). To test the significance of the current model, we simply compare the deviance of the model of interest (D_m) to a model with no explanatory variable (D_0), which is estimated as:

$$D_0 = -2 * \{ n_1 * \ln[P(Y=1)] + n_0 * \ln[P(Y=0)] \}$$

where n_1 is the number of observations with Y=1, n_0 is the number of observations with Y=0, $P(Y=1) = n_1/N$, and $P(Y=0) = n_0/N$ (see Menard 2001). The difference between D_0 and D_m: (D_0 -D_m) is called the *model Chi-square*, denoted by G_m. Under the null hypothesis where all coefficients are equal to 0, G_m follows a Chi-square distribution with k degrees of freedom (k is the number of variables). Therefore, we look at the p-value $P[\chi^2(k) > G_m]$, if it is sufficiently small, (for example p-value < 0.05 for 95% confidence level), then it is essential to reject the null hypothesis and conclude that the coefficients do not simultaneously equal to 0. It is called the *likelihood ratio test*. Sometimes, people test the ratio (G_m/D_0), which is referred as the *likelihood ratio index*. Under the null hypothesis, this ratio approach to 0 (as D_m approach D_0). Typically we look for ratio greater than 0.5 to conclude a strong association between the dependent and independent variables.

It is also important to test the significance of each variable to identify whether it provided sufficient explanation to the model. Likelihood ratio test could be applied in a similar way: we examine the difference between the deviance obtained from the model with the variable of interest and the deviance obtained from the model without the variable, such as: $G = D_{\text{model without the variable of interest}} - D_{\text{model with the variable of interest}}$. Under the null hypothesis, G value follows a Chi-squared distribution with 1 degree of freedom. To test the significance, p-value $P[\chi^2(1) > G]$ is applied as same as before.

Alternatively, a *pseudo t*-test can be applied. For the jth variable of interest, the null hypothesis is $\beta_j = 0$, which implies the conditional probability $P[Y_j=1|X_j]$ does not depend on X_j. When the sample size is large:

$$t_\beta = \frac{\sqrt{N}\beta}{\sigma_\beta} \sim N(0,1)$$

The statistic t_β is called the *pseudo t*-value (see Bierens 2004), which follows the standard normal distribution under the null hypothesis. Again we can take a look at the corresponding p-value: $P[N(0, 1) > t_\beta]$, if it is sufficiently small (for example, p-value < 0.05 for 95% confidence level), then it is essential to reject the null hypothesis and conclude that the corresponding variable is significant. This statistics is available in Matlab statistical toolbox.

4.3 Classification and Regression Tree (CART)

Regression and Classification Tree (CART) is another approach to solve classification type of problems, it is a computerized, non-parametric technique different from traditional statistical methods: CART applies the binary Recursive Partitioning Algorithm (RPA) to best classify samples into a number of non-overlapping regions, each of which correspond to a terminal node in the tree; CART then assign each terminal node into one of the classes based on the criteria of *minimizing expected misclassification costs*.

Assume the sample is consisted of N observations, with N_1 observations belong to class *1* and N_2 observations belong to class *2* (suppose *1* is the bankruptcy class and *2* is the non-bankruptcy class). We denote the prior probability of an observation belong to class *1* as π_1 and the prior probability of an observation belong to class *2* as π_2, the cost of misclassifying a class *1* observation into class *2* is c_{12} and the cost of misclassifying a class *2* observation into class *1* is c_{21}. Consider a terminal node t which has $n_1(t)$ observations from class *1* and $n_2(t)$ observations from class *2*, define the risk of assigning node t to group *1* as:

$$R_1(t) = c_{21}\, p(2, t) = c_{21}\, \pi_2\, p(t|2) = c_{21}\, \pi_2\, n_2(t)/ N_2$$

where $p(2, t)$ is the probability that an observation is from group *2* and falls into node *t*, which equals the prior probability (π_2) multiplied by the conditional probability of a group *2* object falling into node *t* ($p(t|2)$) according to *Bayesian rule*. The conditional probability $p(t|2)$ is calculated by dividing the number of observations of class *2* in node *t* ($n_2(t)$) to the number of observations belong to class *2* (N_2). Intuitively, $R_1(t)$ measures the expected cost of misclassifying class *2* observations into class *1* at node *t*, and similarly, we can calculate the expected cost of misclassifying class *1* observations into class *2* at the same node ($R_2(t)$). The terminal node *t* is then assigned to a class corresponding to the minimum risk, for example, if $R_1(t) < R_2(t)$, assign *t* to class *1*. The risk of the entire tree *T*, denoted by $R(T)$, is the sum of risks of its terminal nodes. A special case exists when misclassification costs $c_{12} = c_{21} = 1$, and the prior probability is the original sample proportion ($\pi_1 = N_1 /N$, $\pi_2 = N_2 /N$), then the risk of assigning node *t* to group *1* is the sample proportion of class *2* observations falling into node *t* ($R_1(t)= n_2(t)/N$), vice versa. In this case, the minimum risk rule implies each terminal node is assigned to a class which has majority representation in this node, and the risk of the tree is simply the overall misclassification rates.

The tree construction process comprises two phases: *growing* and *pruning*. In the growing phase, Recursive Partitioning Algorithm (RPA) is applied to divide the samples into a number of non-overlapping regions. In each iteration, CART seeks the best splitting variable which minimized expected misclassification costs to split the node into two children nodes. The misclassification cost at each node is measured by the *impurity function*. For two-class case, the impurity function is defined as:

$$I(t) = R_1(t)p(1|t) + R_2(t)\, p(2|t)$$

where $p(1|t)$ is the conditional probability that an observation in node *t* is assigned to group *1*:

$$p(1 \mid t) = p(1,t) / p(t) = (\pi_1 n_1(t) / N_1) / (\sum_{k=1}^{2} (\pi_k n_k(t) / N_k))$$

where $p(t)$ is the probability of an observation falling into node *t*. $p(2|t)$ is defined and calculated in the similar way. $I(t)$ can be interpreted as the expected cost of misclassification when observations in node *t* are randomly assigned to two classes. It is not difficult to observe that $I(t) = 0$ only if all observations in *t* belong to the same class, for example, when all observations in node t belong to class 1, then $R_1(t)= p(2|t) =0$. By substituting the expression of $R_1(t)$ and $R_2(t)$:

$$I(t) = (c_{21} + c_{12})\ p(2, t)\ p(1, t) / p(t)= (c_{21} + c_{12})\ p(1|t)\ p(2|t)\ p(t)$$

It can be proved that $I(t)$ attains its maximized value when $p(1|t) = p(2|t) = 0.5$ (see Breiman et al 1984). The impurity of a tree is defined as the sum of impurities of its terminal nodes. Once a terminal node t is further split into two children nodes, t_L and t_R, the decrease in the impurity of the tree is:

$$\Delta I(t) = I(t) - (I(t_L) + I(t_R))$$

where $\Delta I(t)$ is nonnegative and its magnitude depends on the choice of the split. The best split for the current terminal node t expects $(I(t_L) + I(t_R))$ to be minimized, which allowed the split to maximize $\Delta I(t)$. The best splitting variable could be found by an exhaustive search of all possibilities, but better algorithms have been developed to improve the efficiency (see Heping Zhang). The growing phase stopped when one of the following conditions is satisfied: (1) there is only one observation in each child node; (2) no further splitting is able to decrease the impurity of the current tree; (3) an external limit on the depth of the tree is reached.

The pruning phase is aimed at selecting the correct complexity of the tree. The minimal misclassification cost tree obtained from the growing phase usually over-fits the data because it is constructed solely with respect to the samples – not necessarily represent the population characteristics. Pruning is conducted according to the *cost-complexity* criteria. The cost-complexity for each sub-tree of the optimal tree is:

$$CC = R(T) + K^* \text{ (number of terminal nodes of } T\text{)}$$

where $R(T)$ is the sum of misclassification rates for all terminal nodes and K is a nonnegative constant representing a *penalty* for complex trees. The pruning starts from the bottom level, and K increases gradually in the pruning process. The children nodes are pruned away if the resulting change in the predicted misclassification cost is less than K times the change in tree complexity, for example, if $(R(T')- R(T)) < K^*$(number of terminal nodes of T – number of terminal nodes of T'), then the tree T' is preserved as a better solution.

The primary difference between CART and LDA is that the CART classification rule partitions variable space into a number of rectangular regions, while the LDA classification rule partitions the variable space into two half-plane regions. Another difference is the way they deal with prior probabilities and costs of misclassification impacts: LDA models are established by group separation criteria first (such as maximizing inter-group to intra-group variance), and then observations are assigned into corresponding groups based on error costs and prior probabilities; CART, on the other hand, handles variable selection and group assignments simultaneously with the costs and prior probabilities helping to determine the splitting. Therefore, CART is more sensitive to the training data especially when outliers exist. Moreover, the non-parametric property is a desirable feature for analyzing data with unknown distributional properties. Therefore, CART is recommended for the FFS study.

4.4 Chapter Summary

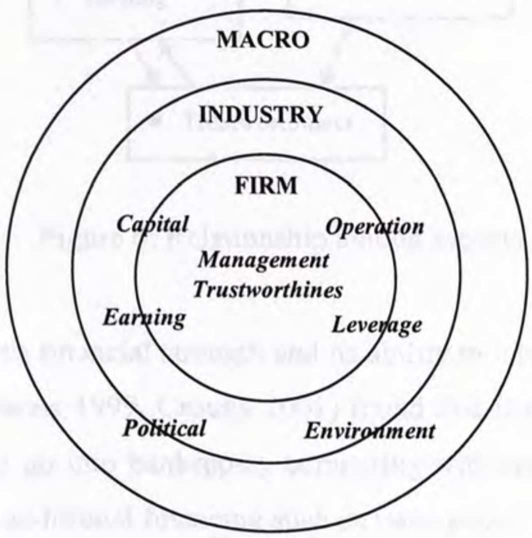
This Chapter discussed theoretical foundations of two statistical techniques: Logit regression and Classification and Regression Tree (CART). The former is a traditional parametric approach widely used in classification problems; the latter is a computerized, non-parametric approach which is relatively new to statistical researchers. We discussed theoretical basis behind the models, and explored characteristics and issues associated with model applications. We recommended Logit regression to develop the credit scoring model and CART for FFS investigations.

CHAPTER 5
PROPOSED STUDY FRAMEWORK

5.1 The COMPLETE Framework

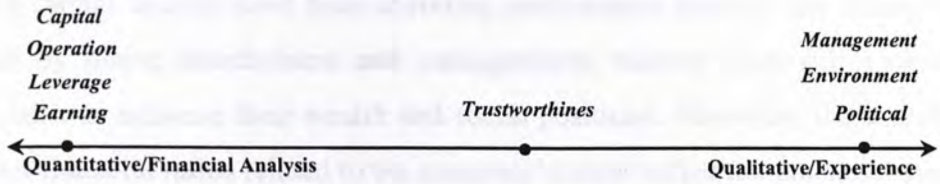
We proposed a conceptual framework – COMPLETE to evaluate credit risks for firms in China. The COMPLETE framework is consisted of 8 aspects – **Capital, Operation, Management, Political, Leverage, Earning, Trustworthiness** and **Environment.** These aspects constitute important considerations in the credit rating process. We can view these aspects from three dimensions:

- Scope: the aspects range from firm/industry-level to macro-economic scope. It complies with rating agencies' considerations, which include earning and cash-flow, quality of assets and the leverage position, industry outlook, the effects of macro-economic events and regulatory changes, etc (Crouhy et al. 2001).



- Figure 4: Scope of aspects

- Measurements: credit rating process includes both quantitative and qualitative analyses. The quantitative analysis is mainly financial analysis based on the firm's financial reports, and the qualitative analysis is concerned with the quality of management, the firm's competitiveness within its industry and its vulnerability to environment changes. **Capital, Operation, Leverage** and **Earning** can be measured quantitatively; **Management, Political** and **Environment** need qualitative analysis; **Trustworthiness** needs analysis from both sides.



• Figure 5: Measurement of aspects

- Relationship: these aspects are inter-related rather than independent. Capital, operation, management, political, leverage, earning and environment are all needed to evaluate trustworthiness of financial information disclosed by the firm, and the trustworthiness consideration is then used to evaluate whether capital, operation, leverage and earning are measured properly.

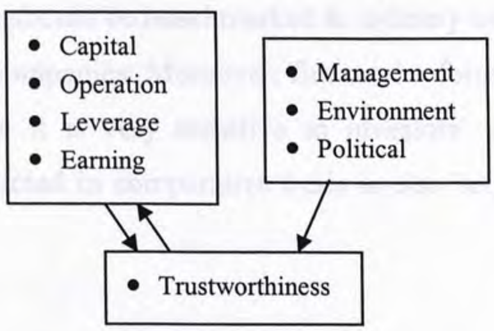


Figure 6: Relationship among aspects

5.1.1 Capital

Capital reflects a firm's financial strength and its ability to meet key term obligations. Previous studies (Persons 1999, Crouhy 2001) found that firms with larger financial size are less likely to go into bankruptcy comparing with small firms, because they have better access to additional financing such as issue public securities. However, all observations in our sample (discussed in Chapter 6) are large companies listed on Shanghai and Shenzhen stock exchange, therefore the analysis is focused on another direction – one should pay special attention to *unused capital*: according to a survey published on *China Securities* in August 2001, the averaged ratio of unused capital (cash + short-term investment) to total assets for 1063 listed companies increased 43.32% comparing with the figure in 2000, among which 54 companies had the proportion of unused capital greater than 50%, and the highest was 87.96% (Xu & Chen 2001). Unused capital is not engaged in the revenue-production process, so it could not bring any profits to the company; moreover, companies with large amount of unused capital usually have poor operating performance because the money were controlled by major shareholders and managements, making them fulfill personal interests such as enhance their wealth and social positions. Therefore, the analysis is focused on financial ratios related to the company's *asset utilization* and *asset quality*, because companies with above phenomenon have lower asset utility.

5.1.2 Operation

Operation evaluates the firm's operating capability, and the analysis is focused on whether the firm has the ability to accumulate capital and grow. Therefore, growth (or potential growth) related financial ratios are employed in this analysis. However, the operation of a firm is affected by its industry condition and macro-economic factors, so the analytic analysis should be benchmarked to industry average and trends, and/or other similar situated companies. Moreover, financial information related is prone to manipulations, because it is very sensitive to investors' decisions. Therefore, an analytic analysis conducted in comparative basis is also helpful in discovering such frauds.

5.1.3 Management

This aspect evaluates the quality of management of a firm. Due to the unique phenomenon – *economic reform* in China, it is necessary to understand features and problems faced by China firms first. China firms can be divided into 3 categories according to the ownership structures: **State-owed Enterprises, Collaborated-owned Enterprises and Township & Village Enterprises,** and **Private-owned Enterprises**. The managerial problems and analysis focuses are presented as follows:

1. **State-owned Enterprises (SOEs):** in an SOE, the State as the controlling shareholder, delegates its power to authorized institutions, and the institutions then send representatives to the shareholders' meeting to exercise shareholders' rights. However, the representatives are usually administrative bureaucrats who lack the incentive and business knowledge to impose effective control on the board and management. On the other hand, the involvement of non-State shareholders, including individuals and legal entities, may have limited influence on the improvement of corporate performance, for example, an entity shareholder

may be another corporation troubled by its own governance issues, and individual shareholders with small shareholdings have few incentives and means to impose significant control on corporate management. Consequently, the external market monitoring force has limited influence, which results in high risk of insider control. In reality, insider control has become the most severe problem that plagues SOEs in China (Yuwa Wei 2001). Therefore, the analysis focus for SOE should be the qualification, responsiveness and personality of boards and management team, especially when such a member is known to be dishonest or corrupted, a warning signal should be raised.

2. **Collaborated-owned Enterprises and Township & Village Enterprises (COEs and TVEs):** the governance problems are mainly caused by factors such as lack of clear laws and lack of standards for corporate behaviors. As the planned-economy dominated China for a long time, corporate development only has a short history; people who participate in corporate practice may not properly understand their roles, their rights and their duties within the corporation. Malfunctions of the shareholders' meeting and the supervisory board are widely reported, abuse of power by directors is not infrequent. Therefore, the analysis should focus on the management process, and one may ask questions such as: does the company have essential control over the accounting process? does it employ qualified staff for essential controls? does it have a proper way to recruit staffs, through just, fair and open channels instead of black-box operations? does it promote a sound corporate culture, such as allowing employees to report improprieties or make suggestions of improvement? Etc. Similar to SOE, the qualification, responsiveness and personality of the management should also be closely examined.

3. **Private-owned enterprises:** these are typically small firms where the ownership is not separated from management, and the decision making is dominated by a single executive or a group of close-related people. Similar to COEs & TVEs, the analysis focus is still the personality of the management, as management approach in China enterprises are always top-down, a significant wrong decision make by the management would lead the firm into bankruptcy. Moreover, unlike SOEs, COEs and TVEs, who can receive certain support from central and/or local

governments, Private-owned enterprises are operated in personal networks, which link strongly but informally with related but legally independent organizations handling key functions such as supply or marketing, therefore they are more vulnerable to financial distress and environmental changes.

As a summary, the agency problem should be considered in the first hand, it included the board of directors and management team. The management process should also be closely examined. In order to facilitate the analysis of the management aspect, it is desirable to develop a profile for each firm, which includes backgrounds of directors and management team members, key decision and fraudulent actions committed by these people (if possible), the firm's governance structure, and significant events (such as litigation and significant losses) happened to the firm. A decision support system should be carefully designed to facilitate information retrieval and cross references.

5.1.4 Political

This aspect examines regulatory trend over an industry or an ownership structure. Over the past 20 years, governmental authority had been decentralized in Chinese enterprises, but the government remains its impact in State-owned Enterprises (SOEs). For example, the government continued to control interest rates and impose credit plans on banks, and consequently, banks extended a large proportion of "policy loans" to loss-making SOEs, while Private-owned enterprises are hard to get loans (Yuwa Wei 2001). Another example is that board members of SOEs are usually government officers, and the business is somewhat favored by *political reasons*, such as Chen Jiulin transformed China Aviation Oil (Singapore) from a small company into an "apotheosis" of oversee Chinese firms just because the holding company is an SOE which granted China Aviation Oil (Singapore) a monopoly position in aviation oil import/export (Caijing 2004). However, along with the opening and development, government policies will be changed and less favor will be granted to SOEs. The political impacts should be examined by experts with proper domain knowledge.

5.1.5 Leverage

Leverage refers to the combination of debts and equity for a company. The analysis is focused on whether the company's combination of debt and equity enables wealth creation without unduly jeopardizing long-term solvency. Highly leveraged firms (those with heavy debts comparing to net worth) lack financial flexibility to respond to crises and/or to take advantages of opportunities; they are more vulnerable to business downturns than lower leveraged firms. However, leverage ratios vary greatly depending on the requirements of particular industry groups. Therefore, analysis should be performed on a comparative basis, which is by comparing the firm to industry averages and other similar situated firms.

5.1.6 Earning

This aspect measures whether the firm is profitable and whether its cash-flow/income is adequate. As earning is the most critical consideration for financial analysis, it is also the most sensitive area for financial statement manipulations. In order to tackle this problem, we also consider the *quality of earning*, which is measured from two perspectives: firstly, we examine the earning-cash-flow correspondence, as firms ultimately collect revenues in cash and pay most expenses with cash, high quality earnings should reflect similar pattern of net income and operating cash-flow, in case these two are going into different directions, something might be wrong; secondly, financial analysts should pay special attention to *contingent liabilities*, which is an off-balance-sheet item. Contingent liabilities are potential financial obligation faced by the firm, such as providing guarantees and facing litigations. These contingencies, once realized, may trigger significant amount of losses. Therefore, analysts need to adjust the earning based on the likelihood of the contingency realization, which requires certain experience and familiarity with the company. In this study, contingency is not considered due to lack of information. Similar to other financial analysis, the analysis should be benchmarked to industry average and trends, and other similar situated companies.

5.1.7 Trustworthiness

Trustworthiness assesses whether financial information disclosed by the firm is reliable. As discussed in Chapter 3, financial statements manipulations are very common and this consideration is indispensable. Credit agencies in the US usually use auditor's opinion as a measurement of information reliability. However, accounting firms in China used to cooperate with their clients and help them to beautify financial reports, and the seriousness can be observed from figures listed at the beginning of Chapter 3. There are two reasons: first, the supervision and punishment towards CPAs are loose; second, CPA firms suffer pressures of adverse reporting from the dominant local industry, and in order to get enough business, they need to do whatever told to.

In order to detect for financial statement fraud, one can analyze the practices and trends in the industry or similar situated companies as compared to the practices and trends of the firm. If the firm is experiencing unusual profits, growth, or significant changes is observed from account balances and key financial ratios, then these should be investigated in detail. Besides, the fundamental side also tells us whether the firm has motivation or characteristics to produce false statements and a number of questions can be considered: does the company have a reliable accounting department and appropriately trained accountants? is internal control sufficient? is the auditing firm independent from the company itself? is management honest? does the company have a troublesome relation? does the company have a sudden change in auditing firm or accounting estimation? Etc. Statistical, neural network and data mining techniques could be integrated into the decision support system to facilitate detection and prediction of financial statement frauds.

5.1.8 Environment

This aspect examines a firm's business networks, and considerations include its major customers, revenue sources, suppliers, and domestic & foreign markets. These considerations provide an image of the vulnerability of the firm to industry conditions and macroeconomic environment changes. Another particular thing to consider is *related parties*. According to the definition in Chinese accounting standard No.1, related parties include parent/subsidiaries, joint ventures, major investors and their controlling companies, and companies controlled by their close relations. Business failure of a related party may have a significant influence on performance of the firm itself. Moreover, if a company has troublesome related parties, the company itself may also have due diligence problem because they are likely to conceal the truth for one another. Here are two cases:

The first case is Zhou Zhengyi (周正毅) and the collapse of his Nongkai Development Group (农凯发展集团). Zhou was in custody for financial scandal and stock price manipulations. His company defaulted HK$ 1.77 billion from the Bank of China (Hong Kong), and at the same time, Nongkai had over 6 billion RMB borrowings from Shanghai banking sectors. Nongkai had extreme complicated investment relations illustrated in Figure 7. Although no official information available, it is believed that Nongkai and its related parties had manipulated financial statements utilizing this relationship, because financial difficulty of a single company can be easily veiled by allocating its debts into related parties.

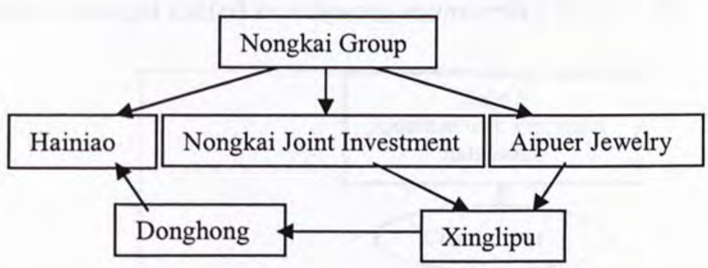


Figure 7: Investment relationship in Nongkai Group

The second case is Guangdong Foshan Nanhaiguanghua Decoration Co. Ltd (广东佛山南海光华板材装饰有限公司), whose owner Fung Mingchang cheated China Industrial & Commercial Bank Nanhai branch for a total amount of RMB 7.421 billion loan. Fung was in custody for creating falsified financial reports and collusion with agents in the bank. He used to hide undesirable financial situations into his 13 related parties. As a non-listed company, the financial manipulation space is even wider.

To facilitate the analysis of business connections, one can add linkage information to company profiles mentioned in 5.1.3, for example, when the analyst click on one company, its business partners and related parties can be appropriately displayed and inter-relationship among board and management team members can be easily retrieved.

5.2 Rating Process

COMPLETE provides a conceptual framework for building up a credit rating system for firms in China. In order to turn this concept into reality, we proposed a 3-stage rating process. The first stage assess the **trustworthiness** aspect; the second stage performs financial analysis on key ratios that reflect **capital, operation, leverage**, and **earning**; the third stage considers qualitative factors, such as **management, political** and **environment** to adjustment for final decision.

The most important concern is how to process the financial data to the surface of truth, so that the output of credit rating model can be meaningful. In the ideal case, the fraudulent actions are *prevented* at the first hand. Therefore, our first proposal brought forward a concept called *continuous monitoring* (Figure 8).

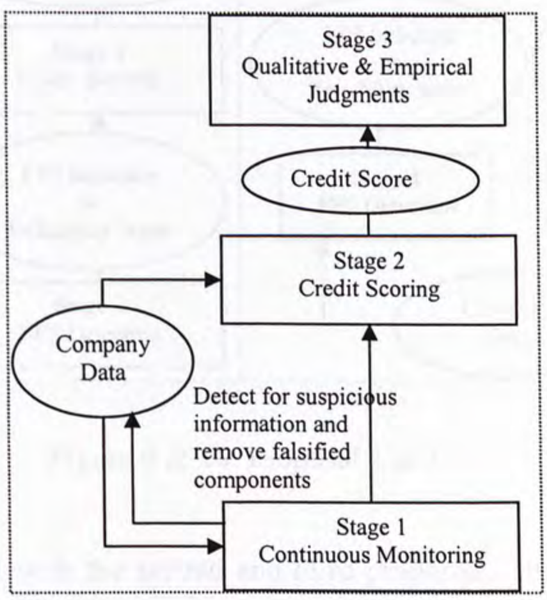


Figure 8: Proposal 1

In the continuous monitoring process, firms are required to update financial statements on an event-driven basis – once a transaction or event happens, the financial report data need to be updated immediately; furthermore, firms should enable credit rating agency to track their accounting process, which implies their accounting system should be opened for inspection at any time. Basically, continuous monitoring integrated *real-time accounting* and *real-time auditing* together, and to some extent, credit agencies also act as external auditors in this process. This idea, if properly implemented, is able to solve due-diligence problem from the bottom line; however, it is very hard to be implemented because it touched the business transactions of individual firms. As every firm is concerned about its business secretes, such information is too confidential to be disclosed. Once implementing this idea, a balance between confidentiality and supervision should be settled, its scope should be clearly defined and it should be strongly imposed by the central government. Therefore, details about continuous monitoring are beyond the scope of this study.

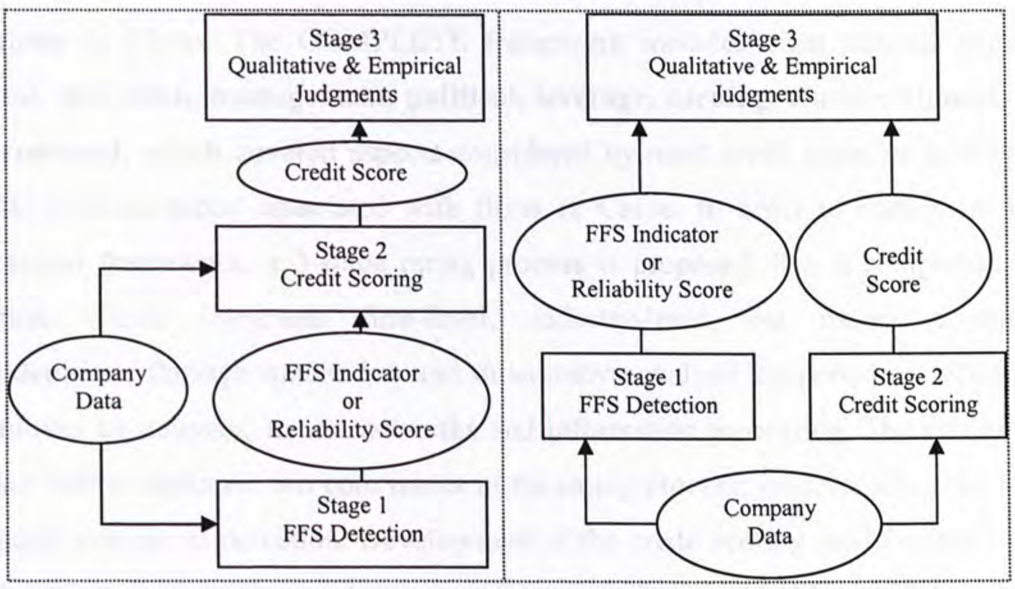


Figure 9 & 10: Proposal 2 & 3

Therefore, we came out with the second and third proposals. These proposals try to build FFS detection models based on public available information, so that the rating company would not touch any sensitive information for individual companies (Figure 9 & 10). Statistical, neural network and data mining techniques can all be applied in building such a model. The output of FFS detection model is a FFS flag or a reliability score, the data type is not a critical issue because multivariate credit scoring models are capable in handling both Boolean and continuous types. However, a score format maybe more desirable because it tells us the likelihood of going into FFS or the degree of confidence. The difference between the two proposals lies in how they treat the trustworthiness aspect. Proposal 2 treat the output of Stage 1 as one of the factors for the credit scoring model, therefore the output of Stage 2 already takes trustworthiness into consideration; proposal 3 simply treat FFS detection and credit scoring as two independent processes, and the outputs are all passed to qualitative and empirical judgments. From bankruptcy risk modeling point of view, proposal 2 is preferred because the model developed is robust to false financial data; From system developer's point of view, proposal 3 maybe more preferred because it provides more flexibility for end users (such as financial analysts) to handle the real situations.

5.3 Chapter Summary

This Chapter discussed the COMPLETE framework proposed to assess credit risks for firms in China. The COMPLETE framework includes eight aspects, namely **capital, operation, management, political, leverage, earning, trustworthiness** and **environment**, which covered aspects considered by most credit agencies as well as special considerations associated with firms in China. In order to implement this conceptual framework, a 3-stage rating process is proposed. It is a comprehensive solution which integrates firm-level, industry-level and macro economic considerations through qualitative and quantitative analysis supported by advanced techniques in statistics, neural networks and information processing. The rest of the Thesis further explored two core issues in the rating process: credit scoring and false financial statements detection. Development of the credit scoring model is discussed in Chapter 6 and false financial statement issues are discussed in Chapter 7.

CHAPTER 6
DEVELOPING THE CREDIT SCORING MODEL

6.1 Overview

Univariate analysis and Logit regression are used in this study. Univariate analysis is used to find out statistically significant indicators to discriminate bankruptcy from non-bankruptcy Chinese firms; Logit regression is employed to develop the credit scoring model. We considered the correlations between explanatory variables because the inclusion of highly correlated variables could result in biased estimation on significant levels of the parameters. This model is further used to investigate severity of false financial data that affect credit scoring result.

6.2 Sample

The sample was obtained from CSMAR annual report database, which collected annual reports of A-share companies listed in Shanghai and Shenzhen exchange covering period 1990 to 2003. This source may not be the most appropriate one for credit risk study, because it only covered listed companies – most of which were transformed from former large State-owned Enterprises (SOEs). However, it is the only data source we could access for the time being. Another difficulty is that CSMAR database does not include information on business failure or bankruptcy, so we defined the bankruptcy condition as: book value of liabilities exceeds book value of assets (or book value of shareholders' equity less than 0). It is only a *technical bankruptcy condition* because some stock codes still exist in the subsequent years. Nevertheless, this technical bankruptcy condition was applied to look for bankruptcy and non-bankruptcy samples. Besides, comparability of the accounting data was considered. China's accounting system and account standards had undergone several major changes since 1993. At present, listed companies follow *Detailed Rules on Information Disclosures for Listed Companies* promulgated by China Securities Regulation Commission (CSRC) in 1999. Therefore a time horizon covering 2000 – 2003 is a better choice because most early data may be not comparable. With all these considerations, 36 firms were identified for the bankruptcy group, and one-year before bankruptcy data are used. We employed the matching principle to look for samples in the non-bankruptcy group. An important benefit of the matching principle is that it removes external factors and the differences can be solely explained by the explanatory variables. The non-bankruptcy group was selected by stratified random basis – matched by total assets, industry and year. Overall, the initial sample is composed of 72 firms, with 36 firms in each group.

6.3 Variables

17 financial ratios were chosen to form the pool of initial variables. These variables are well defined and used to evaluate performance for listed companies in China (中联财务 2003). We further classified these variables into four groups: **capital, operation, liquidity** and **earning** according to the COMPLETE framework

.

	Variable	Definition	Unit
Capital	R1	Revenue / Total Assets	Times
	R2	Gross Profit / Current Assets	Times
	R3	Cost of Good Sold / Inventory	Times
	R4	Gross Profit / Account Receivable	Times
Operation	R5	(Gross Profit / Last Year Gross Profit) - 1	%
	R6	(Shareholders' Equity / Last Year Shareholders' Equity) - 1	%
	R7	Retained Earnings / Shareholders' Equity	%
	R8	(Net Operating Income – Income from Other Operations) / Before Tax Income	%
	R9	(Revenue / Revenue three years ago) ^ (1/3) - 1	%
Leverage	R10	Total Asset / Total Liabilities	%
	R11	Current Asset / Current Liabilities	%
Earning	R12	Net Operating Cash Flow / Current Liabilities	%
	R13	Net Income / Net Worth	%
	R14	Gross Profit / Revenue	%
	R15	Net Operating Income / Total Assets	%
	R16	Net Operating Cash Flow / Net Income	%
	R17	Net Income / Share Capital	%

Table 3: List of Variables

R1 (Revenue / Total Assets) is referred as total asset turnover. This ratio indicates how efficiently a firm generates revenue on each dollar of assets, and it is a common measurement of a firm's asset quality and asset utilization. A larger ratio implies more efficient usage of the assets. **R2 (Gross Profit / Current Assets)** is referred as current asset turn over. It indicates how efficiently a firm generates revenue on each dollar of current assets, and it is another common measurement of a firm's asset quality. **R3 (Cost of Good Sold / Inventory)** is referred as inventory turnover. It measures number of times a firm turn over (or sell) inventory during the year. Generally, a high inventory turnover is an indicator of good inventory utilization and management, and a low turnover may suggest poor capital allocation. **R4 (Gross Profit / Account Receivable)** is referred as account receivable turnover. It measures number of times that trade receivables turnover during the year, the higher the turnover, the shorter the time between sales and collecting cash, and the assets is more liquid.

R5 (Gross Profit / Last Year Gross Profit – 1) is the 1-year growth rate for gross profit. Large growth rate could reflect that a firm has a great deal of development potential. **R6 (Shareholders' Equity / Last Year Shareholders' Equity – 1)** is the 1-year growth rate for shareholders' equity. It measures a firm's ability to accumulate capital and its potential for further expansion. **R7 (Retained Earnings / Shareholders' Equity)** measures the proportion of retained earnings in shareholders' equity, large value implies that a firm is capable to expand on capital generated from its own operations. **R8 ((Net Operating Income – Income from Other Operations) / Before Tax Income)** measures the proportion of income from primary business in before tax income, large value implies that the operating capability is reliable, because a larger proportion of incomes is gained from the primary business rather than some "one-stop" gais. **R9 ((Revenue / Revenue three years ago)^(1/3)-1)** measures 3-year revenue growth of a firm, and reflect its ability for sustainable development.

R10 (Total Assets / Total Liabilities) is the inverse of the leverage ratio. A smaller ratio means the firm has heavy debts and it is more prone to bankruptcy risk. **R11 (Current Asset / Current Liabilities)** is referred as the current ratio or liquidity leverage. A large current ratio indicates a conservative liquidity position, and less chance to go into liquidity problem.

R12 (Net Operating Cash Flow / Current Liabilities) measures how much of a firm's cash profits are available to repay debt. A large ratio indicates that the cash-flow is adequate to cover debts. **R13 (Net Income / Net Worth)** is the referred as return on equity (ROE). It measures the rate of return on the firm's investment in the business, and it is an important indicator on a firm's profitability and earning ability. **R14 (Gross Profit / Revenue)** indicates how much profit is earned on a firm's products without considering of operating and administration costs, and it roughly reflects the firm's cost management and earning ability. **R15 (Net Operating Income / Total Assets**) indicates how efficiently a firm generates operating profits on each dollar of total assets, and measures its operating profitability. **R16 (Net Operating Cash-flow / Net Income)** is the ratio between net operating cash-flow and net income. It is used to check the quality of earning, and a higher ratio indicates higher quality of earning. **R17 (Net Income / Share Capital)** measures net income obtained from each unit of its share capital, and it is another indicator of profitability.

One thing to point out is that the grouping is based on the ratios' definition, interpretation and empirical classification by financial analysts rather than formal statistical analysis (such as factor analysis). Factor analysis is a good tool for evaluating relationship among variables. However, our sample size is not large enough to support factor analysis, so the test is not conducted for the time being.

6.4 Result of the Univariate Analysis

Result of the univariate analysis is presented in Table 4. It indicated that 12 variables are statistically significant at 99% confidence level, and 14 variables are significant at 95% confidence level. The most significant variable is R10 (Total assets / total liabilities, $t = 7.1$), it suggests that firms going into bankruptcy usually have excessive debts. Therefore **leverage** is the most important indicator to analyze bankruptcy risks for firms in China. The second significant variables are R17 (Net income / Share capital, $t = 6.76$) and R15 (Net operating income / total assets, $t = 6.07$), the mean of the bankruptcy group is less than zero, implies that firms going into bankruptcy are unable to generate enough earning from their business. Therefore **earning** is the second most important consideration to analysis bankruptcy risk for firms in China. Other significant variables include R11 (Current ratio, $t = 4.58$), R12 (Cash-flow coverage, $t = 4.23$), R9 (3-year revenue growth, $t = 4.02$), R1 (Total asset turnover, $t = 3.52$), R6 (Shareholders' equity growth, $t = 3.34$) and R7 (Retained earning /

Shareholders' equity, t = 3.31). Therefore, **short-term liquidity**, **sustainability of development**, **asset utilization** and **ability of capital accumulation** are other important considerations in analyzing bankruptcy risk for firms in China.

	Mean		Standard Deviation		
Ratio	**Bankrupt**	**Non-bankrupt**	**Bankrupt**	**Non-bankrupt**	***t*-value**
R1	0.22	0.64	0.22	0.66	**3.52
R2	0.61	1.23	1.26	0.91	*2.35
R3	3.42	12.82	4.93	29.02	1.89
R4	10.73	19.05	35.63	35.47	0.98
R5	-0.15	1.65	0.68	3.93	**2.67
R6	-0.28	1.47	0.39	3.07	**3.34
R7	-5.48	0.20	10.14	0.10	**3.31
R8	0.33	0.58	3.82	1.32	0.37
R9	-0.12	0.43	0.40	0.70	**4.02
R10	1.47	3.02	0.63	1.12	**7.10
R11	0.97	1.75	0.66	0.77	**4.58
R12	-0.01	0.29	0.28	0.32	**4.23
R13	-1.91	0.09	4.09	0.05	**2.88
R14	0.15	0.29	0.26	0.15	**2.64
R15	-0.09	0.06	0.14	0.05	**6.07
R16	-0.50	3.21	1.90	9.33	*2.30
R17	-0.36	0.29	0.54	0.19	**6.76
* Significant at 95% confidence level (t-value >1.96)					
** Significant at 99% confidence level (t-value >2.58)					

Table 4: Univariabe analysis result

Four insignificant variables are R3 (inventory turnover), R4 (account receivable turnover), R8 (net income from primary business over before tax income) and R16 (net operating cash-flow/net income). R3 and R4 were insignificant probably due to business characteristic variation among industries, and R8 was insignificant because of business diversification and investment activities conducted in most listed companies. Overall, the univariate analysis suggested that most 1-year before bankruptcy data are statistically significant to distinguish bankrupt firms from non-bankrupt firms from in China.

6.5 Develop the Bankruptcy Risk Model with Logit Regression

6.5.1 Variable Reduction

As co-linearity among variables will affect accuracy of Logit regression, the first step is aimed at removing highly correlated components. One common method for this purpose is principle component analysis (PCA). However, PCA transforms the variables to another coordinate system, which is not good for this study purpose because newly created coordinate system cannot help explaining relative importance for original variables. Besides, PCA only examines the covariance matrix, which cannot reflect much information for discrimination purpose. Therefore, an alternative approach is employed: firstly, the correlation among the variables is examined (see appendix). As R4 and R8 are two most insignificant variables, those strongly correlated ($p<0.05$) with R4 and R8 are first eliminated. After this step, the remaining variables are R1, R5, R6, R9, R11, R13, R14 and R16. It is also observed that R5, R6 and R14 are strongly correlated with R9, as R9 has the highest *t*-value, therefore, R5, R6 and R14 are eliminated. As a result, five variables – R1, R9, R11, R13 and R16 are used in Logit regression. Note that two variables with highest *t*-value – R10 and R15 did not enter the model. In fact, inclusion of a value with very high *t*-value would probably mask the importance of other variables and undermine the overall explanation power of the model developed. This variable deduction scheme finally lead to at least one variable cover each category, for example, R1 – Capital, R9 – Operation, R11 – Leverage and R13 & R16 – Earning.

6.5.2 The Logit Model

The Logit model developed is presented in Table 5. We can observe that the *p*-value for each parameter is smaller than 0.05, indicating variables selected are statistically significant at 95% confidence level. The model Chi-squared test has a very small *p*-value ($p < 0.000$) and the likelihood ratio index is 0.83, indicating a strong relationship between the dependent and independent variables.

As the sampling for bankrupt and non-bankrupt firms is one-to-one (36:36), it is expected that the probability of bankruptcy is 50%. Therefore, it is reasonable to assume a cut-off point of 0.5. By re-substituting the dataset into the model, we get a hit ratio of 88.89% for bankrupt firms and 91.67% for non-bankrupt firms. The overall hit ratio is (88.89%+91.67%) / 2 = 90.28%.

	Coefficient	Standard Error	*p*-value
Constant	4.2789	0.2370	0.0000
R1	-3.7353	1.3932	0.0073
R9	-2.7564	0.1709	0.0000
R11	-1.4483	0.4758	0.0023
R13	-4.6583	0.0073	0.0000
R16	-0.7745	0.1046	0.0000

Deviance: 23.8151 (p<0.000), R_L^2: 0.83
Correctly Predicted:
Non-bankrupt: 91.67%
Bankrupt: 88.89%
Overall: 90.28%

Table 5: Result of the Logit regression

Recall that Logit regression fits linear logistic model for binary response variable using maximum likelihood estimation, positive coefficients indicate increasing probability of bankruptcy with corresponding explanatory variables, and vice versa. The coefficients for all independent variables are negative, indicating the variables have negative impacts on bankruptcy (larger value indicates the probability of bankruptcy is small). It complied with the variable explanations in 6.3.

The ouput can be directly interpreted as probability of bankruptcy. However, in credit scoring, people prefer that higher score indicates healthier of the firm, therefore, following transformation is made:

$$score = P[Y_j = 0 \mid X_{1,j} \dots X_{k,j}] = 1 - P[Y_j = 1 \mid X_{1,j} \dots X_{k,j}] = (1 - fit\_value) * 100$$

and the output score can be interpreted as probability of non-bankrupt (in percentage).

6.5.3 Finding the Optimal Cut-off Point

In order to find out the optimal cut-off point, a trial-and-error process is conducted. The cut-off scores were tried from 5 to 100 (incremented by 5 each time), and prediction accuracy was calculated in each trial (Table 6). With maximizing over-all hit ratio as the criteria, a cut-off score of 60 is more desirable. Bankruptcy correctly predicted is 94% and non-bankruptcy correctly predicted is 92%, overall prediction accuracy is (94%+92%)/2=93%. We can observe from the univariate analysis that the bankruptcy group had larger variance, so the optimal cut-off point is a little bit shifted to the right.

Cutoff Point	# Correctly Predicted		% Correctly Predicted		
	Bankrupt	Non-Bankrupt	Bankrupt	Non-Bankrupt	Total
5	24.00	36.00	0.67	1.00	0.83
10	27.00	36.00	0.75	1.00	0.88
15	28.00	36.00	0.78	1.00	0.89
20	29.00	36.00	0.81	1.00	0.90
25	30.00	35.00	0.83	0.97	0.90
30	30.00	35.00	0.83	0.97	0.90
35	32.00	35.00	0.89	0.97	0.93
40	32.00	34.00	0.89	0.94	0.92
45	32.00	34.00	0.89	0.94	0.92
50	32.00	33.00	0.89	0.92	0.90
55	32.00	33.00	0.89	0.92	0.90
60	34.00	33.00	0.94	0.92	0.93
65	34.00	33.00	0.94	0.92	0.93
70	34.00	32.00	0.94	0.89	0.92
75	35.00	31.00	0.97	0.86	0.92
80	35.00	30.00	0.97	0.83	0.90
85	36.00	29.00	1.00	0.81	0.90
90	36.00	26.00	1.00	0.72	0.86
95	36.00	20.00	1.00	0.56	0.78
100	36.00	0.00	1.00	0.00	0.50

Table 6: Finding optimal cut-off point

6.5.4 Justify the Result from Probit Regression

In order to examine reliability of above results, Probit regression is applied on the same dataset. Probit regression is similar to Logit except the underlying probability function is assumed to be the cumulative normal distribution function. Result of Probit regression is presented in Table 7. The *p*-value for each parameter is less than 0.05, indicating that variables selected for Logit model are also significant in the Probit model at 95% confidence level. The model Chi-squared test has a very small *p*-value ($p < 0.000$) and the likelihood ratio index is 0.80, indicating a strong relationship between the dependent and independent variables. Again, apply 0.5 as the cutoff point and conduct the same back testing procedures, Probit gives a hit ratio of 91.67% for both bankrupt and non-bankrupt firms. The hit ratio for bankrupt group is slightly higher than Logit, and the hit ratio for non-bankrupt group is the same. Overall, the performances are quite similar.

	Coefficient	Standard Error	*p*-value
Constant	2.5496	0.6918	0.0002
R1	-2.5666	1.0728	0.0167
R9	-1.3039	0.5612	0.0202
R11	-0.8336	0.3174	0.0086
R13	-0.5085	0.0230	0.0000
R16	-0.4069	0.0083	0.0000
Deviance: 28.8218 ($p < 0.000$), R_L^2: 0.80 Correctly Predicted: Non-bankrupt: 91.67% Bankrupt: 91.67% Overall: 91.67%			

Table 7: Result of Probit regression

6.6 Chapter Summary

This Chapter described the credit scoring model we developed to assess credit risk for firms in China. The sample consisted 72 A-share companies listed in Shanghai and Shenzhen exchang. 17 financial ratios were included in the initial pool of variables. Univariate analysis is used to look for key indicators reflecting bankruptcy risks, and the result indicated that 14 financial ratios are good enough to distinguish bankruptcy from non-bankruptcy firms with one-year before bankruptcy data, and variables related to leverage, earnings, short-term liquidity, sustainability of development, asset utilization and ability of capital accumulation could provide helpful suggestions for credit analysts. A credit scoring model is developed with Logit regression, which achieved an overall prediction accuracy of 90.28%. This model is applied as a benchmark in our false financial statements investigations presented in Chapter 7.

CHAPTER 7
INVESTIGATING FALSE FINANCIAL STATEMENTS

7.1 Overview

One of the difficulties in rating a firm in China is due to heavy financial statements manipulations, which significantly distorted their performance. This Chapter is aimed at making in depth analysis over this phenomenon and providing solutions. The first part of this Chapter looked into a well-known financial scandal case – Hubei Lantian Co. Ltd. (蓝田股份造假案), illustrated the impact of FFS on credit scoring, and demonstrated how to apply the COMPLETE framework to handle this case. The second part employed Classification and Regression Tree (CART) method to analyze 22 false financial statements and some interesting findings were discussed.

7.2 Impact of False Financial Statements (FFS) on Credit Risk Assessments – Evidence from Lantian's Case

Hubei Lantian Co. Ltd was listed on Shanghai stock exchange in 1996, with major business in livestock, aquaculture, and beverage production. Since 1997, Lantian's performance was unbelievably good: its total assets expended more than 10 times from 1996 to 2000 (266 million to 2838 million), revenue grew 4 times (460 million to 1840 million) and net income increased more than 7 times (59 million to 432 million). It was recognized as a "flag of China's agriculture industrialization". However, its performance declined quickly in year 2001, and until that time, people realized the "outstanding" performance was just a result of "creative" accounting.

It is said the Chinese firms keep two sets of financial statements – one for the investors and one for themselves. Lantian was just this case. Once the financial scandal was discovered, Lantian was forced to disclose the true figures and large discrepancies could be observed (Table 8). Three areas were significantly overstated: cash and cash-flow (3669.90% and 130.37% overstated); fixed assets (200.65% overstated); and performance related accounts (revenue 4732.45% overstated, operating income 720.12% overstated, net income 4138.98% overstated).

Major Accounts (RMB) & Financial Ratios	Falsified	True	Percentage Overstated
Cash	167144107	4433648	3669.90
Account Receivable	28907068	301837050	(90.42)
Inventory	236384087	124494782	89.87
Current Assets	433106704	431825347	0.30
Fixed Assets	2169016000	721436338	200.65
Total Assets	2837651898	1155472868	145.58
Current Liabilities	560713384	955275978	(41.30)
Total Liabilities	657633752	1038574778	(36.68)
Revenue	1840909605	38094775	4732.45
Operating Income	503349653	-81169602	720.12
Net Income	431628612	-10686569	4138.98
Retained Earnings	1004801104	-585955496	271.48
Cash-flow from Operations (net)	785829628	-64060513*	1326.70
R1 (Total Assets Turnover)	0.65	0.03	1865.76
R9 (3-year Revenue Growth)	0.04	0.16	(75.00)
R11 (Current Ratio)	0.77	0.45	70.88
R13 (ROE)	0.20	-0.09	316.63
R16 (Cash-flow Coverage)	1.82	-5.99	130.37
* This figure is estimated from other financial accounts.			

Table 8: Major Accounts & Financial Ratios of Lantian Co. Ltd. 2000

Five financial ratios in our credit scoring model calculated from the original data were 0.65, 0.04, 0.77, 0.20 and 1.82 respectively, the credit score obtained was 84.6, indicating a pretty healthy financial situation; however, the adjusted data gave those values of 0.03, 0.16, 0.45, -0.09 and -5.99 respectively, and the credit score was merely 0.03, which definitely indicating the bankruptcy situation.

7.3 Evaluating the Trustworthiness Aspect for Lantian

This Section demonstrated how to apply the COMPLETE framework to evaluate trustworthiness of financial data disclosed by Lantian, and discussions were divided into two parts: the first part compared Lantian's financial accounts and financial ratios to industry average and other similar situated firms, and the second part looked for warning signals and possible motivations for financial scandal.

7.3.1 Financial Accounts and Financial Ratios

According to Lantian's annual report in 2000, aquaculture and beverage were two major revenue sources, where 69% of its revenue came from aquaculture and 29%

came from beverage (Table 9). Therefore, these two sectors were compared against corresponding industry average and similar situated firms.

Unit (RMB)	Revenue	Costs	Gross Profit
Aquaculture	1270467421	874848361	395619060
Beverage	529243070	290235934	239007136
Others	41199114	40782400	416714
Total	1840909605	1205866695	635042910

Table 9: Segment Report of Lantian Co. Ltd. 2000

Lantian's revenue from aquaculture was 1270 million RMB, 3 times higher than industry average. According to Lantian's website, its aquaculture base was 200,000 mou, therefore its revenue per mou was about 6350 RMB. Comparing to a similar situated firm – Wuchangyu Co. Ltd., whose revenue per mou was about 1723 RMB, Lantian's performance was 3 times higher. However, both firms produced similar aquatic products and their aquaculture bases were both in Hubei province, therefore, such performance deviation should be taken cautiously.

	Lantian		**Industry Avg.**		**Dongting Aqua-culture**	**Wu-chang-yu**	**LoLo**
	Aqu.	Bev.	Aqu.	Bev.			
Revenue (million RMB)	1270	529	272	798	74	112	821
Gross Margin (%)	31.1	45.2	33.1	33.7	54.8	47.7	28.2
Acc. Rec. (million RMB)	29		101	133	14	12	13
Current Ratio	0.77		3.71	2.40	9.66	3.76	3.23
Inv. to Current Assets (%)	64.5		8.8	29.5	2.0	2.7	15.6
Fixed Assets to Total Assets (%)	76.4		22.6	31.5	7.9	10.3	46.7

Table 10: Comparison of major accounts & financial ratios

Lantian's profit margin on beverage was 45.2%, about 36% higher than industry average. Chengde LoLo Co. Ltd. – one of the largest almond juice producers in China, only had a profit margin of 28.2%. Lantian's beverage products were lotus juice and mineral water, the production system should be similar to LoLo; it was also noticed that the brands of Lantian – Yelian juice and Yeou juice, were far less famous than LoLo, therefore its high profit margin was also questionable.

By looking into other financial ratios, more questions came out: percentage of account receivable in revenue was merely 2.3%, 7 times lower than Dongting Aquaculture Co. Ltd. (18.9%) and 3 times lower than Wuchangyu Co. Ltd. (10.7%), both of which operated in aquaculture industry in the neighboring districts. Furthermore, Lantian's current ratio in 2000 was 0.77, 3 times lower than the aquaculture industry average and 2 times lower than the beverage industry average. This number also implied its current assets were unable to cover current liabilities. From 1998 to 1999, its current ratio was strictly declining, implied its performance got worse than ever. These facts became contradictions to Lantian's outstanding revenue and high profit margin, which implied the performance was overstated.

Year	Fixed Assets Growth (%)	Operating Income Growth (%)	Current Ratio
1998	136.84%	156.38%	1.68
1999	103.27%	42.50%	1.07
2000	27.73%	-19.18%	0.77

Table 11: Lantian's Performance 1998 – 2000

Lantian's fixed assets to total assets ratio in 2000 was 76.4%, significantly higher than industry averages (aquaculture – 22.6%, beverage – 31.5%) and other similar situated companies (Dongting – 7.9%, Wuchangyu – 10.3%, LoLo – 46.7%). As large deviations from industry average and similar situated firms were observed, the account was possibly under manipulations. Lantian's fixed assets growth was 137% in 1998 and 103% in 1999, comparing with the growth rates in operating income, one can infer that two accounts were manipulated together.

7.3.2 Warning Signals

From Lantian's annual report, it is found that another company – China Lantian (Holding) Co. Ltd. had close business relation with it, for example, China Lantian had provided more than 222 million capitals to Hubei Lantian through related party transactions. From agri.com.cn (金农网), it is found that the legal representative – Qu

Zhaoyu (瞿兆玉) of China Lantian also held the post of director & general manager in Hubei Lantian, and he had been the chairman of Hubei Lantian from 1996 to 1998. These facts were signals of insider-control.

As the entry barrier for agriculture industry was relatively low, and the Chinese government provided favored policies on agriculture firms (such as lower tax), lots of speculators were engaged in this industry and FFS had been a common phenomenon. The performance of the whole industry was not good since 1998. These facts further provided incentives and possibilities for Lantian to conduct financial manipulations.

By combining both quantitative analysis and qualitative analysis, one could conclude that Lantian's financial data cannot be trusted. Fixed assets, revenue and incomes were significantly overstated.

7.4 Analyze FFS with Statistical Tools

The previous Section discussed a number of analytical tests for detecting Lantian's financial scandal. However, such a process is highly dependent on experience and subjective judgments. Therefore it is necessary to employ a scientific approach to look for common features about financial scandals in Chinese firms, and apply the findings to facilitate FFS detection. In this Section, we introduce ten most serious financial scandal cases among listed companies in China. Their manipulation histories are presented shortly as follows:

1. Guangxia Yinchuan Industry Co. Ltd (stock code 000557, abv. Guangxia) made use of fabricated contracts and reports created a large number of fictitious transactions. Fabricated revenue was 37% in 1999, and 508% in 2000.

2. Hubei Lantian Co. Ltd (stock code 600709, abv. Lantian) made a series of FFS from 1996 to 2000. In year 2000 alone, the fabricated fixed assets and revenue were RMB 1.4 billion and 1.8 billion respectively, which were twice and forty times higher than corresponding true figures.

3. Zhang Jiajie Tourism Development Co. Ltd (stock code 000430, abv. Zhang Jiajie) applied inappropriate revenue recognition methods and manipulated its revenues and incomes from 1996 to 1998. Overstated revenue was RMB 80 million in 1996 and 43 million in 1997. Overstated net income was 5.28 million in 1998.

4. Shenyang Dawn Garments Co. Ltd (stock code 600167, abv. Dawn Garments) manipulated its financial statements in 1999 by "magnifying" its business: assets, liabilities, shareholders' equities, revenue and before tax incomes were overstated by RMB 90 million, 20 million, 74 million, 150 million and 87 million respectively. The original 34 million losses were turned into 52 million profits.

5. Zhengzhou Baiwen Co. Ltd Group (stock code 6008998, abv. Zhengzhou Baiwen) used accounting tricks such as creating fictitious revenue, understating liabilities and adjusting timing of revenue recognition to enhance its performance. From 1996 to 1997, total fabricated income was RMB 144 million.

6. Dongfang Boiler Group Co. Ltd (stock code 600786, abv. Dongfang Boiler) shifted 176 million revenue and 38 million incomes of year 1996 to 1997, and 226 million revenue and 47 million incomes of year 1997 were shifted to 1998. These actions created a steady growth of net assets turnover from 1996 to 1998.

7. Luoyang Chundu Foodstuffs Company Limited (stock code 000885, abv. Luoyang Chundu) failed to disclose several important related party transactions (related to RMB 431 million) in its 1999 financial reports; it also failed to record 380 million expenses, which either affected its asset valuation or profits.

8. Luzhou Laojiao Co. Ltd (stock code 000568, abv. Luzhou Laojiao) issued a readjustment announcement in May 2003, in which its performance differed a lot from that reported in 2002 annual report: net income decreased from 51 million to 31 million, EPS decreased from 0.0978 to 0.0588, and return on net assets decreased from 3.457 percent to 2.077 percent.

9. Chengdu Hongguang Industrial Co. Ltd (stock code 600083, abv. Hongguang) applied financial manipulation tricks such as creating fictitious sales and

inventory and turned 103 million losses in 1996 into 157 million profits. In 1997, the reported loss amount was 32 million less than the true losses.

10. Hainan Dadonghai Tourism Centre Holdings Co. Ltd (stock code 000613, abv. Dadonghai) used tricks such as inappropriate accounting estimates for account receivables and creating fictitious revenue repeatedly from 1993 to 1997. The total amount of fabricated income was as high as RMB 228 million.

The primary objective of making FFS was boosting revenue and net incomes, because firms making losses would receive punishments from the regulatory body and unable to get enough funding accordingly. Most firms above boost their revenue by creating fictitious transactions. Overstating revenue would jointly raise income, assets and shareholders' equity. Sometimes firms understate expenses, and the same effects can be achieved. FFS could also be produced through manipulating accounts that can be estimated from different accounting methods, such as inventories and fixed assets. A flowchart of FFS production is illustrated in Figure 11.

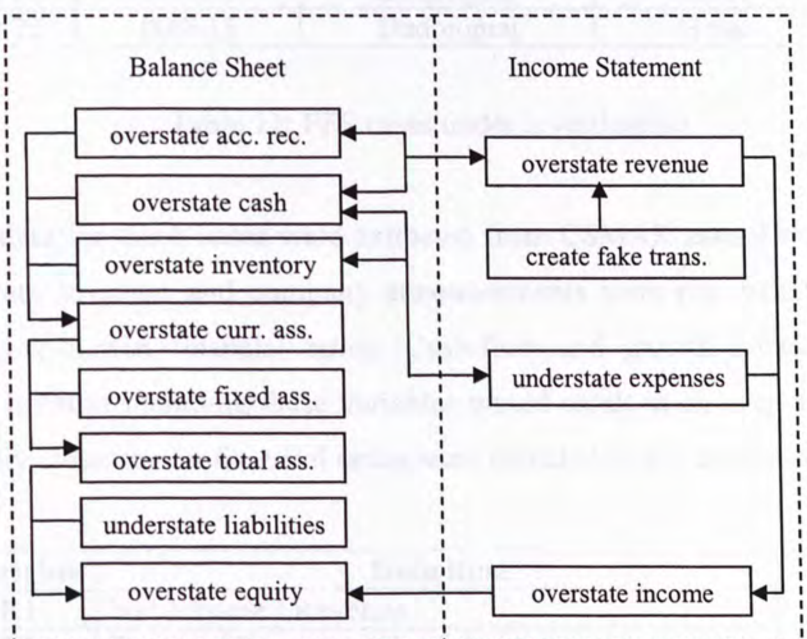


Figure 11: Flowchart of FFS production

7.4.1 Data Acquisition

Twenty-two FFS cases under investigation were summarized in Table 12 (financial data from 1993 – 1995 were not available for the last case).

Case	Ticker Code	Company Abv.	Industry	Year
Case 1	000557	Guangxia	Agriculture	1999
Case 2	000557	Guangxia	Agriculture	2000
Case 3	600709	Lantian	Agriculture	1996
Case 4	600709	Lantian	Agriculture	1997
Case 5	600709	Lantian	Agriculture	1998
Case 6	600709	Lantian	Agriculture	1999
Case 7	600709	Lantian	Agriculture	2000
Case 8	000430	Zhang Jiajie	Tourism	1996
Case 9	000430	Zhang Jiajie	Tourism	1997
Case 10	000430	Zhang Jiajie	Tourism	1998
Case 11	600167	Dawn Garments	Commerce	1999
Case 12	600898	Zhengzhou Baiwen	Retail	1996
Case 13	600898	Zhengzhou Baiwen	Retail	1997
Case 14	600786	Dongfang Boiler	Machinery	1996
Case 15	600786	Dongfang Boiler	Machinery	1997
Case 16	600786	Dongfang Boiler	Machinery	1998
Case 17	000885	Luoyang Chundu	Food	1999
Case 18	000568	Luzhou Laojiao	Beverage	2002
Case 19	600083	Hongguang	Electrics	1996
Case 20	600083	Hongguang	Electrics	1997
Case 21	000613	Dadonghai	Hotel	1996
Case 22	000613	Dadonghai	Hotel	1997

Table 12: FFS cases under investigation

Financial data for these cases were extracted from CSMAR annual report database. As corporate structure and company announcements were not included, we could only apply common financial ratios. Cash-flow and growth related ratios were excluded, because including these variables would result in an even smaller sample size. Finally, 28 common financial ratios were included in this analysis (Table 13).

Variable	Definition	Unit
R1	Net Income / Revenue	%
R2	Retained Earnings / Total Assets	%
R3	Cost of Good Sold / Inventory	%
R4	Quick Assets / Current Liabilities	%
R5	Financial Expenses / Revenue	%
R6	EBIT / Total Assets	%
R7	Revenue / Total Assets	%
R8	Net income / Total Assets	%
R9	Operating Income / Total Assets	%
R10	(Operating Income + Financial Expenses) / Total	%

	Assets	
R11	Operating Income / Shareholders' Equity	%
R12	Net Income / Shareholders' Equity	%
R13	Operating Expenses / Revenue	%
R14	Operating Income / Revenue	%
R15	12 * Fixed Assets / Revenue	month
R16	12 * Current Assets / Revenue	month
R17	12 * (Note Receivable + Account Receivable) / Revenue	month
R18	12 * Inventory / Revenue	month
R19	12 * (Note Payable + Account Payable) / Revenue	month
R20	Current Assets / Current Liabilities	%
R21	12 * Cash / Revenue	month
R22	12 * Quick Assets / Selling and Administrative Expenses	month
R23	12 * Cash / Selling and Administrative Expenses	month
R24	(Current Assets + Fixed Assets) / Shareholders' Equity	%
R25	Shareholders' Equity / Total Assets	%
R26	Fixed Assets / Shareholders' Equity	%
R27	12 * Current Liabilities / Revenue	month
R28	12 * Shareholders' Equity / Revenue	month

Table 13: Variables used in constructing the FFS detection model

7.4.2 Univariate analysis

Univariate analysis is aimed at identifying significantly distorted financial ratios in these FFS cases. Two experiments were conducted: the first experiment compared financial ratios for each FFS firm with corresponding industrial average. The industrial average was calculated by averaging the financial ratio for all available firms in that industry in corresponding year (excluding the FFS firms). The second experiment compared each FFS firm with a similar situated non-FFS firm for the whole FFS period, and similarity is measured by financial size (measured by book value of assets), performance (measured by revenue) and industry.

The result of univariate analysis by comparing FFS with industry average is illustrated in Table 14. The analysis indicated that FFS is associated with significantly higher retained earning to total assets (R2, significant at 90% confidence level) and higher quick assets to selling and administrative expenses (R22, significant at 95%

confidence level). It is most likely a result from revenue and income manipulations, where retained earnings and quick assets were boost by fictitious transactions.

	mean		standard deviation		
Ratio	**FFS**	**non_FFS**	**FFS**	**non_FFS**	**t-value**
R1	0.14	0.10	0.24	0.14	-0.57
R2	0.09	0.03	0.10	0.04	**-2.49
R3	1.33	1.68	0.80	0.94	1.34
R4	0.04	0.03	0.05	0.03	-0.73
R5	0.21	0.19	0.18	0.12	-0.41
R6	0.08	0.06	0.08	0.04	-1.10
R7	0.61	0.52	0.50	0.21	-0.76
R8	0.07	0.05	0.07	0.03	-1.08
R9	0.07	0.05	0.08	0.03	-1.43
R10	0.09	0.06	0.08	0.03	-1.57
R11	0.13	0.38	0.13	2.35	0.50
R12	0.12	0.39	0.12	2.33	0.55
R13	0.05	0.08	0.09	0.08	1.28
R14	0.14	0.09	0.24	0.14	-0.80
R15	10.85	12.21	11.92	7.27	0.46
R16	17.59	17.72	10.68	4.96	0.05
R17	4.81	3.72	4.28	1.92	-1.09
R18	4.99	3.55	4.34	1.74	-1.44
R19	2.04	1.64	2.17	0.87	-0.80
R20	1.82	2.09	0.87	0.93	1.01
R21	0.29	0.39	0.28	0.23	1.35
R22	12.82	8.69	9.47	3.53	*-1.91
R23	3.84	2.93	4.08	1.87	-0.95
R24	2.28	3.29	1.85	8.59	0.54
R25	0.55	0.58	0.23	0.15	0.56
R26	0.61	0.66	0.45	1.12	0.23
R27	10.63	11.85	8.03	3.41	0.65
R28	18.95	20.90	16.40	10.08	0.47
*Significant at 90% confidence level (t-value > 1.72) **Significant at 95% confidence level (t-value > 2.07) ***Significant at 99% confidence level (t-value > 2.82)					

Table 14: Univariate analysis result for FFS comparing with industrial average

As contrast to observations from Western studies, financial ratios of sample FFS firms in China were distorted such that better than average performance was observed (higher retained earning to total assets, higher quick assets to selling and administrative expenses). Therefore "good performing companies" should be taken more cautiously, for example, if some figures appeared to be too good, investors or auditors should immediately raise the question – is it possible? However, as the sample size is small, it is not sufficient to generalize this conclusion to all FFS cases in China. The other variables seemed to be not statistically significant, which is probably due to high variance and/or non-uniform patterns exist in the FFS group.

The result of univariate analysis by comparing FFS with industry peers is illustrated in Table 15. The analysis indicated that FFS is associated with significantly higher cost of good sold to inventory (R3, significant at 90% confidence level), higher quick ratio (R4, significant at 95% confidence level), higher financial expense to revenue (R5, significant at 90% confidence level), lower operating expense to revenue (R13, significant at 90% confidence level), higher current assets to revenue (R16, significant at 90% confidence level), higher receivables to revenue (R17, significant at 95% confidence level), higher current ratio (R20, significant at 95% confidence level), higher quick assets to selling and administrative expenses (R22, significant at 99% confidence level), and higher cash to selling and administrative expenses (R23, significant at 95% confidence level).

	Mean		**standard deviation**		
Ratio	**FFS**	**non_FFS**	**FFS**	**non_FFS**	**t-value**
R1	0.14	0.14	0.24	0.10	0.12
R2	0.09	0.07	0.10	0.05	-0.92
R3	1.33	0.95	0.80	0.45	*-1.92
R4	0.04	0.01	0.05	0.04	**-2.30
R5	0.21	0.12	0.18	0.10	*-2.03
R6	0.08	0.09	0.08	0.08	0.34
R7	0.61	0.58	0.50	0.40	-0.17
R8	0.07	0.08	0.07	0.06	0.36
R9	0.07	0.08	0.08	0.07	0.13
R10	0.09	0.08	0.08	0.07	-0.21
R11	0.13	0.15	0.13	0.16	0.59
R12	0.12	0.15	0.12	0.13	0.77
R13	0.05	0.12	0.09	0.16	*1.92
R14	0.14	0.14	0.24	0.11	0.04
R15	10.85	13.02	11.92	9.09	0.68
R16	17.59	12.85	10.68	6.77	*-1.76
R17	4.81	2.57	4.28	2.94	**-2.03
R18	4.99	3.86	4.34	2.60	-1.04
R19	2.04	1.49	2.17	1.30	-1.01
R20	1.82	1.38	0.87	0.35	**-2.20
R21	0.29	0.25	0.28	0.28	-0.44
R22	12.82	5.37	9.47	3.67	***-3.44

R23	3.84	1.41	4.08	1.13	**-2.69
R24	2.28	1.78	1.85	0.64	-1.20
R25	0.55	0.56	0.23	0.15	0.17
R26	0.61	0.76	0.45	0.25	1.43
R27	10.63	9.23	8.03	3.95	-0.74
R28	18.95	16.28	16.40	9.12	-0.67
*Significant at 90% confidence level (t-value > 1.72) **Significant at 95% confidence level (t-value > 2.07) ***Significant at 99% confidence level (t-value > 2.82)					

Table 15: Univariate analysis result of FFS comparing with industry peer

It is not difficult to observe that univariate analysis with industry peer is more effective, as more FFS indicators can be explored. Similar to the previous observation, most financial ratios of these cases were distorted such that better performance was observed (higher cost of good sold to inventory, higher quick ratio, lower operating expenses to revenue, higher current ratio, higher quick assets to selling and administrative expenses and higher cash to selling and administrative expenses). Besides, nearly all the financial ratios were associated with accounts (such as revenue, cash, account receivables and inventory) affected by transactions in the revenue cycle, others were indirectly affected by these accounts (for example, overstate account receivables will boost current assets).

Two experiments in the univariate analysis provided 11 significantly distorted financial ratios from sample FFS firms. It is found that the manipulations were conducted so aggressively that the performance of these firms appeared better than their industry peers and/or industry average. It is also observed that accounts related to income (revenue and expenses) and accounts affected by transactions in the revenue cycle (revenue, cash, account receivables and inventory) were most sensitive for manipulations. However, univariate analysis is unable to provide a boundary or threshold to discriminate FFS from non-FFS, therefore multivariate strategy is also necessary. Due to small sample size and possibly non-normal distribution of the samples, a non-parametric approach is preferred. CART is selected as the multivariate methodology, and the analysis result is discussed in the next Section.

7.4.3 Result of the CART Analysis

The CART analysis is aimed at finding out the boundaries or thresholds to discriminate FFS from non-FFS. Once the classifications were obtained, we also want to explore the relationship between severity of financial statements manipulation and the actual financial and operating status of the company. Table 16 summarized the true figures of four key financial accounts (RMB) and percentage overstated in FFS. The true figures were obtained from readjustment announcements of these firms disclosed in later years. However, not every FFS firm had such disclosure and only 12 cases had corresponding information.

Stock Code (year)	Revenue (percentage overstated)	Net Income (percentage overstated)	Total Assets (percentage overstated)	Shareholder's Equity (percentage overstated)
000557 (99)	383579946.11 (37.14)	127786600.85 (0.00)	2191828506.7 (10.86)	942460049.35 (0.00)
000557 (00)	130261227.08 (507.82)	-135823810.33 (407.49)	1429597831.2 (120.43)	-340353948.53 (455.08)
600709 (00)	38094774.28 (4732.45)	-10686569.22 (4238.98)	1155472867.8 (145.58)	129253932.45 (1585.38)
600709 (99)	24238787.13 (7538.29)	-22879728.64 (2342.28)	874766383.2 (167.27)	268102569.94 (551.54)
000430 (98)	51881314.46 (57.56)	-3026424.63 (971.37)	382818289.09 (-21.5)	194558190.8 (12.46)
000430 (97)	74637404.09 (0.00)	16803355.77 (25.54)	255765602.19 (1.68)	187774926.74 (2.29)
600167 (99)	256654941.65 (59.52)	-22092554.43 (260.27)	722999232.86 (13.09)	486935263.07 (15.22)
600786 (98)	930774189.34 (0.00)	-12615680.52 (114.36)	1736772532.62 (3.14)	486447331.89 (11.02)
600786 (97)	1099603666.53 (0.00)	11272484.23 (131.20)	1859447299.39 (0.80)	519975695.46 (2.84)
000885 (99)	455777719.99 (0.00)	11900880.05 (102.34)	956159003.71 (1.27)	622859690.49 (1.96)
000568 (02)	1044804516.22 (0.00)	30561457.09 (1.10)	2466806230.09 (0.00)	1497344371.69 (0.02)
600083 (97)	270659389.93 (0.00)	-229221127.45 (13.45)	1798121329.88 (5.98)	611509264.47 (18.41)

Table 16: True financial performance and percentage manipulated

Two experiments were conducted, which compared FFS with industry average and industry peers, data used in univariate analysis were applied here again. The result of

CART analysis by comparing FFS with industry average is illustrated in Figure 12. The depth of the tree is four and number of leaf nodes (represented by rectangles) is five. Four leaf nodes correspond to the FFS class and one corresponds to the non-FFS class. The figure (x:y) at each node represents the number of FFS cases to non-FFS cases, for example, (22:22) at the root node means there are 22 FFS cases and 22 non-FFS. We can observe that all the leaf nodes are "pure" (observations in each node belong to only one class), so the misclassification cost for this tree is 0. The cost printed in each internal node represents the misclassification cost of the sub-tree obtained by pruning associated leaf nodes away, for example, if we prune away those two leaf nodes at the bottom level, then the sub-tree obtained (with four leaf nodes) has a misclassification cost of 0.02.

The CART analysis provided four indicators – operating expenses to revenue (R13), revenue to total assets (R7), 12 * current liabilities / revenue (R27), and 12 * inventory / revenue (R18). It also indicated that FFS were associated with operating expenses to revenue less than 0.018 and/or revenue to total assets less than 0.261 and/or 12 * current liabilities / revenue less than 7.822 and/or 12 * inventory / revenue less than 7.567.

Operating expense to revenue (R13) is recognized as the most effective discriminator, and a firm with it less than 1.8 percent (<0.018) is classified as FFS. The majority of FFS cases (14 out of 22) are identified by this single indicator, including:

- Case 2: Guangxia 00
- Case 3-7: Lantian 96-00
- Case 8-10: Zhang Jiajie 96-98
- Case 12-13: Zheng Baiwen 96-97
- Case 14-16: Dongfang Boiler 96-98

Revenue to total assets (R7) is recognized as the second effective discriminator, and a firm with it less than 0.26 is classified as FFS. 5 of out the rest 8 FFS companies fell into this case:

- Case 1: Guangxia 99
- Case 19-20: Chengdu Hongguang 96-97

• Case 21-22: Hainan Dadonghai 96-97

12 * current liabilities / revenue (R27) is recognized as the third effective discriminator, and a firm with it less than 7.82 is classified as FFS. 2 of out the rest 3 FFS companies fell into this case:

• Case 11: Dawn Garments 99
• Case 17: Luoyang Chundu 99

12 * inventory / revenue (R18) is recognized as the last effective discriminator, and a firm with it larger than 7.56 is classified as FFS. The last FFS case is identified:

• Case 18: Luzhou Laojiao

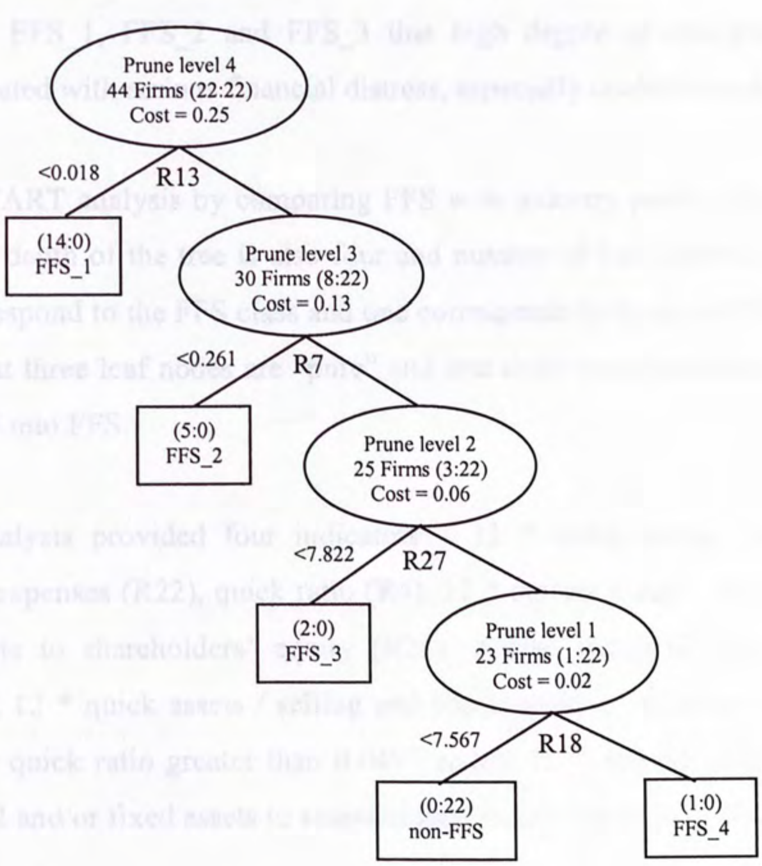


Figure 12: Output of CART of FFS comparing with industrial average

It is observed that most extreme FFS cases went into node FFS_1, including:

- Guangxia 00: percentage overstated for net income and shareholders' equity were 407% and 455% respectively, the 136 million loss were turned into 417 million profit, and 340 million negative equity were turned into positive.
- Lantian 99 & 00: percentage of revenue overstated was as high as 4732% in 1999 and 7538% in 2000; the dangling business was reported to be highly profitable.
- Zhang Jiajie 98: net income overstated was 971%, which turned 22 million loss into 35 million profit

It implied that aggressive revenue manipulations would result in a lower than average operating expense to revenue ratio. Cases fall into FFS_2 and FFS_3 also had high percentage of overstated revenue and income, but not as aggressive as those in FFS_1. The single case contained in FFS_4 did not distort major financial data a lot. It is observed from FFS_1, FFS_2 and FFS_3 that high degree of manipulation was typically associated with serious financial distress, especially unable to make profits.

The result of CART analysis by comparing FFS with industry peers is illustrated in Figure 13. The depth of the tree is also four and number of leaf nodes is five. Four leaf nodes correspond to the FFS class and one corresponds to the non-FFS class. We can observe that three leaf nodes are "pure" and one node contains misclassification of one non-FFS into FFS.

The CART analysis provided four indicators – 12 * quick assets / selling and administrative expenses (R22), quick ratio (R4), 12 * current assets / revenue (R16), and fixed assets to shareholders' equity (R26). It also indicated that FFS were associated with 12 * quick assets / selling and administrative expenses greater than 13.2895 and/or quick ratio greater than 0.0497 and/or 12 * current assets / revenue less than 4.8982 and/or fixed assets to shareholders' equity less than 0.4821.

12 * quick assets / selling and administrative expenses (R22) is recognized as the most effective discriminator, and a firm with it greater than 13.2895 is classified as FFS. 10 out of 22 FFS cases are identified by this single indicator, including:

- Case 1-2: Guangxia 99-00
- Case 3-4: Lantian 96-97

- Case 8-10: Zhang Jiajie 96-98
- Case 12-13: Zheng Baiwen 96-97
- Case 17: Luoyang Chundu

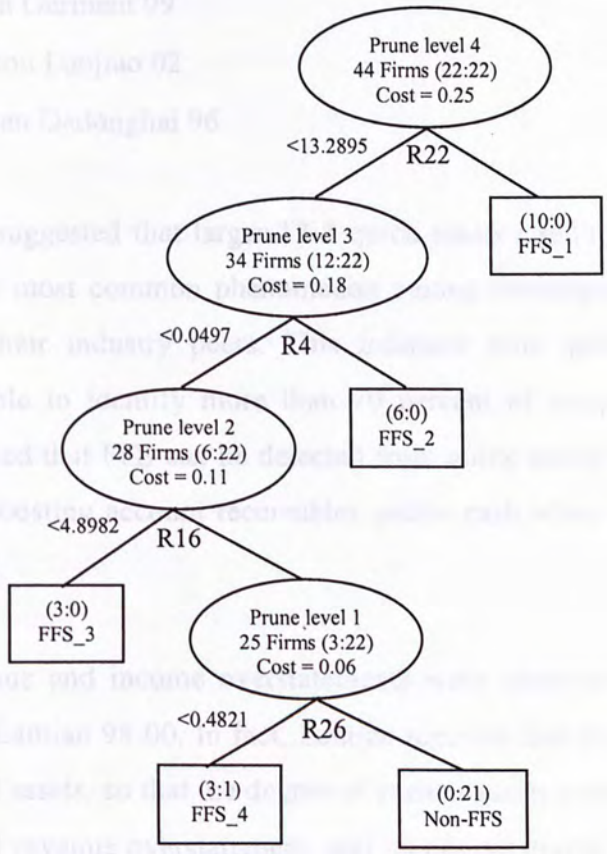


Figure 13: Output of CART of FFS comparing with matched firms

Quick ratio (R4) is recognized as the second effective discriminator, a firm with it greater than 4.97 percent (>0.0497) is classified as FFS. 6 of out the rest 12 FFS companies fell into this case:

- Case 14-16: Dongfang Boiler 96-98
- Case 19-20: Chengdu Hongguang 96-97
- Case 22: Hainan Dadonghai 97

12 * current assets / revenue (R16) is recognized as the third effective discriminator, and a firm with it less than 4.8982 is classified as FFS. 3 of out the rest 6 FFS companies fell into this case:

- Case 5-7: Lantian 98-00

Fixed assets to shareholders' equity (R26) is recognized as the last effective discriminator, and a firm with it less than 0.4821 is classified as FFS. The remaining 3 FFS cases were correctly identified:

- Case 11: Dawn Garment 99
- Case 18: Luzhou Laojiao 02
- Case 21: Hainan Dadonghai 96

This experiment suggested that larger 12 * quick assets / selling and administrative expenses was the most common phenomenon among investigated FFS cases when compared with their industry peers. This indicator plus quick assets to current liabilities, was able to identify more than 70 percent of sample FFS cases. Both indicators suggested that FFS can be detected from quick assets, which is probably a consequence of boosting account receivables and/or cash when revenue and income were manipulated.

The largest revenue and income overstatements were observed from node FFS_3, which contained Lantian 98-00. In fact, Lantian reported that the incomes were used to invest for fixed assets, so that the degree of current assets overstatement was much lower than that of revenue overstatement, and its current assets to revenue ratio was lower. Node FFS_4 contained Luzhou Laojiao, which did not distort major financial data significantly and Dawn Garment, which made FFS in the most comprehensive way. Their distinctions from non-FFS peers were not very obvious and the discriminator selected by CART made a little bit misclassification at this node.

7.4.4 Discussion on Analysis Results

The first finding in our analysis is that financial ratios of sample FFS firms in China were distorted such that these firms appeared to be performing better than average or other similar firms in its industry. It is a contrast to observations from Western studies. Furthermore, companies with serious financial distress were associated with aggressive manipulations, especially high degree of revenue and income overstatement made by fabricated transactions. Therefore, financial statements from such "good performing companies" should be analyzed even more cautiously, because banks were usually willing to do business with such good performing companies, for example, granting loans. Failing to identify the existence of such data manipulation or fabricated transactions would lead to significant losses.

It is also observed that FFS of most firms in this analysis were made on a consecutive basis, where false figures were derived from the old false figures. The evidence can be found from CART analysis, where most FFS cases done by the same company were grouped into the same node. This finding suggested that horizontal comparison would be more effective than vertical comparison in discovering such fraud. Because in the latter case, the financial accounts were examined against previous years' accounts of the same company, once FFS were made on a consecutive basis, it was difficult to observe unusual or significant changes. Horizontal comparison, on the other hand, compares financial accounts and trends of the firm with industry average or other similar situated firms, the fraud would be easier to detect. An efficient FFS detection system should facilitate horizontal analysis with comparable industry peers and industry dependent variables dynamically.

Another important observation is that severity of FFS was industry dependent, for example, observed FFS came from several traditional industries where entry barriers were relatively low, and regulations were not tight. For industries with higher entry barriers or higher degree of internationalization, FFS was rarely observed. The main reason is that the investors, either corporate or professional, would impose higher requirements on the quality of financial data so that they can make comparison with other investment candidates. However, in other industries, which are more localized or traditional, the quantity of reliable data might be insufficient and it was also difficult to develop reliable industry average to support horizontal analysis.

7.5 Chapter Summary

This Chapter first investigated the influence of falsified financial data on credit score from Hubei Lantian Co. Ltd.'s financial scandal case. This case indicated that financial statement manipulation would able to totally change a company's credit worthiness. Then, we demonstrated how to apply the COMPLETE framework to evaluate the trustworthiness aspect through a number of analytical tests. Furthermore, we applied univariate and multivariate statistical tools to analyze several well-known financial scandal cases in China. In order to reach a satisfactory depth, only cases with well investigated evidence are used, and we did not use a large sample size. Univariate *t*-test was applied to compare common financial ratios of these FFS with corresponding industry average and industry peer, 11 common financial ratios were found significantly distorted in these FFS cases. This result would be helpful to lay the foundation for further FFS research. Classification and Regression Tree (CART) analysis was applied with the objective to explore the relationship between severity of financial statements manipulation and the actual financial and operating status of the company. It is found that companies with serious financial distress were associated with aggressive manipulations, especially high degree of revenue and income overstatement, which was able to disguise a nearly bankruptcy company into a highly profitable one. It is also found that FFS of most investigated companies were made on a consecutive basis, which suggested that horizontal analysis would be better than vertical analysis for detecting the fraud.

A more complete study on FFS should be conducted with a larger sample size and explanatory variables covering a wider range (this study did not include variables related to account trends, corporate governance and auditors). Besides, the integration of FFS detection with credit scoring is not conducted, because we cannot identify enough samples with both characteristics of FFS and bankruptcy. One remedy is to release the bankruptcy requirement to business failure, which can be defined as companies receiving ST/PT from CSRC. Many of these companies had a sudden change in performance, and the reports before the changes were suspected for manipulations. This study will be conducted in the future.

CHAPTER 8
SUMMARY

This research is aimed at developing a reliable credit rating system for firms in China. Credit rating in general is an opinion of creditworthiness of the obligator. The rating process is complicated in terms of scope and evaluation. Both quantitative and qualitative evaluations should be applied, and the conclusion is highly relied on subjective judgments. However, previous research seldom address due diligence issues, which cannot be ignored when rating a firm in China, otherwise, the result would be quite unreliable.

The major contribution of this research is the COMPLETE frame we proposed, which integrated due diligence considerations into the credit rating process and provided a guideline to evaluate credit risks for firms in China. We proposed three prototypes of rating process, the first one, which included **continuous monitoring** of the accounting process, is a suggestion for regulatory authorities to strengthen the regulatory enforcements and prevent financial statement fraud at the bottom line. The latter two proposals aim at developing quantitative models based on public data to evaluate the trustworthiness aspect, which is a more practical way for credit agents to handle due diligence problems.

Multivariate statistical techniques are employed in this study to find key indicators and develop the credit scoring model. The sample is consisted of 72 A-share companies listed in Shanghai and Shenzhen exchange. The result of the univariate analysis indicated that 14 financial ratios are good enough to distinguish bankruptcy from non-bankruptcy firms with one-year before bankruptcy data, and variables related to leverage, earnings, short-term liquidity, sustainability of development, asset utilization and ability of capital accumulation could provide helpful suggestions for credit analysts. A credit scoring model is developed with Logit regression, which achieved an overall prediction accuracy of 90.28%.

The impact of falsified financial statements is then investigated. We combined both case study approach and quantitative approach in order to reach an in depth analysis.

Firstly, we demonstrated how to apply the COMPLETE framework to evaluate the trustworthiness aspect of one large financial scandal case in China and pointed out that financial statement manipulation is able to totally distort a company's credit worthiness. Furthermore, we applied univariate and multivariate statistical tools to analyze several well-known financial scandal cases. Univariate *t*-test was applied to compare common financial ratios of these FFS with corresponding industry average and industry peer, 11 common financial ratios were found significantly distorted in these FFS cases. This result would be helpful to lay the foundation for further FFS researches. Classification and Regression Tree (CART) analysis was applied with the objective to explore the relationship between severity of financial statements manipulation and the actual financial and operating status of the company. It is found that companies with serious financial distress were associated with aggressive manipulations, especially high degree of revenue and income overstatement, which was able to disguise a nearly bankruptcy company into a highly profitable one. It is also found that FFS of most investigated companies were made on a consecutive basis, which suggested that horizontal analysis would be better than vertical analysis for detecting the fraud. The future research is aimed at developing an FFS detection model, and integrating it with the credit scoring model. More data sources have to be identified, and a decision support system is to be completed.

REFERENCES

AICPA, "Consideration of Fraud in a Financial Statement Audit", Statement on Auditing Standards No. 82, American Institute of Certified Public Accountants, NY, 1997

Altman Edward I. Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy, Journal of Finance, Vol.23, No.4, pp589-609, 1968

Altman Edward I. The Success of Business Failure Prediction Models, Journal of Banking and Finance, Vol.8, pp171-198, 1984

Altman Edward I. Predicting Financial Distress of Companies: Revisiting the Z-Score and ZETA Models, 2000

Altman Edward I., Marco G. Varetto F. "Corporate distress diagnosis: Comparisons using linear discriminant analysis and neural networks", Journal of Banking and Finance, Vol.18, pp505-529, 1994

Altman Edward I., Haldeman R., Narayaman P. "ZETA analysis: A New Model to Identify Bankruptcy Risk of Corporations", Journal of Banking and Finance, pp29-54, 1977

Amir E. Atiya "Bankruptcy Prediction for Credit Risk Using Neural Networks: A Survey and New Results", IEEE Transactions on Neural Networks, Vol.12, No.4, pp929-935, 2001

Dobson A. J. "An Introduction to Generalized Linear Models", Chapman & Hall, 1990

Back B., Teija Laitinen, Kaisa Sere, Michiel van Wezel "Choosing Bankruptcy Predictors Using Discriminant Analysis, Logit Analysis and Genetic Algorithms", Technical Report No. 40, Turku Center for Computer Science, 1996

Beasley M. "An empirical analysis of the relation between board of director composition and financial statement fraud", The Accounting Review, Vol.71, No.4, pp443-466, 1996

Beasley S.M., Carcello J.V. and Hermanson D.R. "Fraudulent Financial Reporting: An Analysis of US Public Companies", Research Report, COSO, 1999

Beaver W. "Financial ratios as Predictors of Failure, Journal of Accounting Research", Vol.5, pp71-111, 1966

Beaver W. "Alternative Accounting Measures as Predictors of Failure, The Accounting Review", Vol.43, No.1, pp113-122, 1968

Breiman L, Friedman J.H., Olshen R.A. and Stone C.J., "Classification and Regression Trees", Wadsworth International Group, 1984

CBRC, "Promoting the reform, opening-up and sound development of the banking sector through supervision by law", Delivered at the first China International Service Industries Convention and Expo, June 30, 2004

CreditMetrics, Technical Document, JP Morgan, 1997

Credit Suisse, "CreidtRisk+: A Credit Risk Management Framework", Credit Suisse Financial Products, 1997

Crouhy M, Galai D, Rober Mark "Prototype risk rating system, Journal of Banking and Finance", Vol.25, pp47-95, 2001

Crouhy M, Galai D, Rober Mark "A comparative analysis of current credit risk models, Journal of Banking and Finance", Vol.24, pp59-117, 2000

Dallas E. Johnson "Applied Multivariate Methods for Data Analsysts", Duxbury Press, 1998

Deakin, E.B. "A discriminant analysis of predictors of business failure", Journal of Accounting Research, Vol.10, pp167-179, 1972

Desai, V.S., J.N. Conway, G.A. Overstreet Jr. "A Comparison of Neural Networks and Linear Scoring Models in the Credit Union Environment", European Journal of Operational Research, Vol.95, pp24-37, 1996

Dobson A.J. "An Introduction to Generalized Linear Models", Chapman & Hall, 1990

Edmister, R. "An Empirical Test of Financial Ratio Analysis for Small Business Failure Prediction", Journal of Financial and Quantitative Analysis, Vol.7, No.2, pp1477-1493, 1972

Eisenbeis, R.A. "Pifalls in the application of discriminant analysis in business, finance and economics", Journal of Finance, Vol.22, No.3, pp975-890, 1977

Eisenbeis, R.A. "Problems in applying discriminant analysis in credit scoring models", Journal of Banking and Finance, Vol.2, pp205-219, 1978

El Hennawy R., Morris R. "The significance of base year in developing failure prediction models", Journal of Business Finance and Accounting, pp209-223, 1983

Fanning K.M. and Cogger K.O. "Neural Network Detection of Management Fraud Using Published Financial Data", International Journal of Intelligent Systems in Accounting, Finance & Management, Vol.7, No.1, pp21-41, 1998

Frydman, H., Altman, E. and Kao, D. "Introducing Recursive Partitioning for Financial Classification: The Case of Financial Distress", Journal of Finance, Vol.40, No.1, pp269-291, 1985

Heping Zhang, "Recursive Partitioning and Tree-based Methods", http://www.quantlet.com/mdstat/scripts/csa/html/node208.html

Hosmer D.W. and Lemeshow Stanley, "Applied Logistic Regression", Wiley, 1989

Hamer, M. "Failure prediction: Sensitivity of classification accuracy to alternative statistical methods and variable sets", Journal of Accounting and Public Policy, Vol.2, No.4, pp289-307, 1983

Jae H. Min, Youngchan Lee "A Practical Approach to Credit Scoring", Proceedings of ICEB, Beijing, China, December, 2004

Jerry W. Lin, Mark I. Hwang and Jack D. Becker "A fuzzy neural network for assessing the risk of fraudulent financial reporting", Managerial Auditing Journal, Vol.18, No.8, pp657-665, 2003

Lennox C. "Identifying Failing Companies: A Re-evaluation of the Logit, Probit and DA Approaches, Journal of Economics and Business", Vol.51, pp347-364, 1999

Lipsher L.E. "The New Chinese Accounting Standards, Regulation & Enforcement", Lipsher Accountancy Corporation based in the PRC, May, 2002

Liu Mingkang, "The letter from Liu Ming Kang, the Chairman of CBRC, to Mr Jaime Caruana, Chairman of the Basel Committee on Basel II", July 31, 2003

McLeay S.S. "Student's *t* and the distribution of financial ratios", Journal of Business Finance and Accounting Vol. 13, No. 2, pp209-222, 1986

Menard Scott, "Applied Logistic Regression Analysis", 2nd edition, Sage Publications, 2001

Merton R.C. "On the Pricing of Corporate Debt: The Risk Structure of Interest Rates", Journal of Finance, Vol. 29, No. 2, pp449-470, 1974

Morris, R. "Early Warning Indicators of Corporate Failure: A Critical Review of Previous Research and Further Empirical Evidence", Ashgate Publishing in association with the Institutte of Chartered Accountants in England and Wales. 1997

Ohlson J.A. "Financial Ratios and the Probabilitiec Prediction of Bankruptcy", Journal of Accounting Research, Vol.18, No.1, pp109-131, 1980

Persons O.S. "Using Financial Information to Differentiatie Failed vs. Surviving Finance Companies in Thailand: An Implication for Emerging Economies", Multinational Finance Journal, Vol.3, No.2, pp127-145, 1999)

Quinlan J.R. "Comparing connectionist and symbolic learning methods", Proceedings of a workshop on Computational learning theory and natural learning systems, Vol. 1, pp445-456, 1993
Richard J. Bolton and David J. Hand "Statistical Fraud Detection: A Review", Statistical Science, Vol.17, No.3, pp235-255, 2002

Schilit H.M. "Financial Shenanigans", New York: McGraw-Hill, Inc., 1993

Shirata C.Y. "Financial Ratios as Predictors of Bankruptcy in Japan: An Empirical Research", Asian-Pacific Interdisciplinary Research in Accounting (APIRA), Osaka, Japan, 1998

Spathis C.T. "Detecting false financial statements using publised data: some evidence from Greece", Managerial Auditing Journal, Vol.17, No.4, pp179-191, 2002

Tam Kar Yan and Kiang Melody Y. "Managerial Applications of Neural Networks: The Case of Bank Failure Predictions", Management Science, Vol.38, No.7, pp926-947, 1992

West D. "Nerual Network Credit Scoring Models", Journal of Computers and Operations Research, Vol.27, pp1131-1152, 2000

Wilcox, J.W. "A prediction of business failure using accounting data", Journal of Accounting Research, Vol.11, pp163-179, 1973

Yuwa Wei "Comparative corporate governance", Kluwer Law International, 2003

Zavgren, C.V. "Assessing the vulnerability to failure of American industrial firms: a logistic analysis", Journal of Business Finance and Accounting, Vol.12, pp19-45, 1985

Zimmermann, H.J. "Fussy Sets, Decision Making, and Expert Systems", Boston: Kluver Academic Publisher, 1987

中国上市公司业绩评价报告, 中联财务顾问有限公司中国上市公司业绩评价课题, 2003

刘姝威, 应立即停止对蓝田股份发放贷款, 2001-10-26

吴革, 财务报告陷阱，北京出版社出版集团文津出版社，2004

周永亮, 中国企业前沿问题报告，中国社会科学出版社，2001

徐念榕，陈晓翔, 募股资金在手，使用效率堪忧，中国证券报，2001-8-31

邱宜干, 我国上市公司会计信息披露问题研究，江西人民出版社，2003

金德环, 李胜利, 如何进行年报解读:企业财务比率综合分析, 上海证券报, 2004-5-21

飞草, 上市公司十大管理舞弊案分析及侦查研究, http://www.e521.com/cpa/al/000024.htm

飞草，四类上市公司业绩基本不可信，http://www.homeway.com.cn2001.9.26

神话是如何凋零？蓝田股份业绩之谜曝光, CCTV 经济半小时，2004-7-5

张帆，王晓冰，李箐，傅凯文，成败陈久霖，财经，2004-12-13

Appendix: correlation of variables

	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12	R13	R14	R15	R16	R17
R1	1	0.76	0.44	0.32	0.17	-0.23	0.07	-0.05	0.17	-0.07	-0.06	0.09	0.32	-0.18	0.68	-0.26	0.31
		(0.00)	*(0.01)*	*(0.06)*	*(0.33)*	*(0.18)*	*(0.68)*	*(0.76)*	*(0.31)*	*(0.69)*	*(0.74)*	*(0.61)*	*(0.06)*	*(0.30)*	*(0.00)*	*(0.13)*	*(0.07)*
R2		1	0.45	0.55	0.08	-0.39	0.12	0.02	0	0.12	-0.22	0.42	0.29	-0.09	0.55	-0.11	0.45
			(0.01)	*(0.00)*	*(0.66)*	*(0.02)*	*(0.47)*	*(0.93)*	*(1.00)*	*(0.50)*	*(0.19)*	*(0.01)*	*(0.09)*	*(0.61)*	*(0.00)*	*(0.52)*	*(0.01)*
R3			1	0.67	0.05	-0.09	-0.04	0.01	0.16	-0.05	-0.27	0.28	0.14	0.02	0.54	-0.12	0.28
				(0.00)	*(0.76)*	*(0.60)*	*(0.82)*	*(0.94)*	*(0.34)*	*(0.78)*	*(0.11)*	*(0.10)*	*(0.41)*	*(0.91)*	*(0.00)*	*(0.48)*	*(0.09)*
R4				1	-0.01	-0.09	0.01	0.03	0	0.36	-0.05	0.68	0.05	0.09	0.45	-0.01	0.41
					(0.96)	*(0.59)*	*(0.97)*	*(0.86)*	*(0.98)*	*(0.03)*	*(0.76)*	*(0.00)*	*(0.75)*	*(0.61)*	*(0.01)*	*(0.97)*	*(0.01)*
R5					1	0.17	-0.06	-0.04	0.47	-0.11	-0.29	-0.11	0	-0.16	0.1	-0.12	0.12
						(0.31)	*(0.73)*	*(0.82)*	*(0.00)*	*(0.51)*	*(0.09)*	*(0.50)*	*(1.00)*	*(0.36)*	*(0.55)*	*(0.49)*	*(0.48)*
R6						1	-0.09	-0.11	0.53	-0.04	0.02	-0.2	-0.09	0.43	0	-0.09	-0.13
							(0.62)	*(0.52)*	*(0.00)*	*(0.84)*	*(0.93)*	*(0.25)*	*(0.61)*	*(0.01)*	*(0.98)*	*(0.62)*	*(0.45)*
R7							1	0.45	0.01	0.38	0.18	0.06	0.64	0.11	0.09	0.04	0.21
								(0.01)	*(0.97)*	*(0.02)*	*(0.30)*	*(0.72)*	*(0.00)*	*(0.51)*	*(0.60)*	*(0.81)*	*(0.21)*
R8								1	-0.03	0.48	0.03	0.01	-0.05	-0.12	-0.12	-0.28	-0.18
									(0.84)	*(0.00)*	*(0.85)*	*(0.93)*	*(0.75)*	*(0.50)*	*(0.49)*	*(0.09)*	*(0.29)*
R9									1	-0.14	-0.16	-0.19	0.08	0.34	0.19	-0.18	0.04
										(0.43)	*(0.36)*	*(0.27)*	*(0.66)*	*(0.04)*	*(0.27)*	*(0.29)*	*(0.81)*
R10										1	0.54	0.61	-0.07	0.13	0.08	-0.02	0.07
											(0.00)	*(0.00)*	*(0.69)*	*(0.44)*	*(0.66)*	*(0.93)*	*(0.68)*
R11											1	0.15	0	0.09	0.03	-0.03	0.01
												(0.37)	*(0.99)*	*(0.59)*	*(0.87)*	*(0.85)*	*(0.97)*
R12												1	0.05	0.18	0.24	0.16	0.37
													(0.77)	*(0.30)*	*(0.15)*	*(0.36)*	*(0.03)*
R13													1	0.08	0.38	-0.02	0.58
														(0.66)	*(0.02)*	*(0.91)*	*(0.00)*
R14														1	0.18	0.07	0.19
															(0.31)	*(0.68)*	*(0.28)*
R15															1	-0.11	0.67
																(0.52)	*(0.00)*
R16																1	0.06
																	(0.75)
R17																	1